

JLIS.it

Italian Journal of Library and Information Science

Rivista italiana di biblioteconomia, archivistica e scienza dell'informazione

Pubblicazione periodica semestrale (esce in giugno e in dicembre)

ISSN: 2038-5366 (print) – ISSN: 2038-1026 (online)

Website: <http://jlis.it> – Email: info@jlis.it

ABBONAMENTO 2012:

Italia € 50,00; Estero € 60,00.

SUBSCRIPTION 2012:

Italy € 50,00; Overseas € 60,00.

Ledizioni – LediPublishing – Via Alamanni 11

20141 - Milano - Italia

Tel. +39-0245071824 – Fax +39-0242108107 – IVA/VAT: IT04627080965



Vol. 3, n. 1 (Giugno/June 2012)

JLIS.it

Italian Journal of Library and Information Science
Rivista italiana di biblioteconomia,
archivistica e scienza dell'informazione

Università di Firenze
Dipartimento di Scienze dell'Antichità, Medioevo e
Rinascimento e Linguistica

Direttore = Editor in chief

Mauro Guerrini (Università di Firenze)

Condirettore = Co-editor

Gianfranco Crupi (Sapienza Università di Roma)

Managing editor

Andrea Marchitelli (CILEA)

Vicedirettori = Associate editors

Andrea Capaccioni (Università di Perugia),

Graziano Ruffini (Università di Firenze)

Staff

Elena Bortot (layout), Valentina Demontis, Ilaria Fava,

Giulia Manzotti

JLIS.it è edita dall'Università di Firenze, Dipartimento di Scienze dell'Antichità, Medioevo e Rinascimento e Linguistica.

PROPOSTE DI PUBBLICAZIONE & PEER-REVIEW: Le submission, compiute tramite il sito web della rivista, verranno inizialmente esaminate da un editor e, superata la prima valutazione, saranno inviate a due revisori per il processo di peer review, al termine del quale verrà notificata l'accettazione o meno del contributo, o l'eventuale richiesta di modifiche.

DIRITTI: JLIS.it applica una licenza "Creative Commons – Attribuzione" (CC-BY) a tutto il materiale pubblicato.

JLIS.it is published by the University of Florence, Department of studies on the Antiquities, Middle Age, the Renaissance and Linguistics.

SUBMISSION & PEER-REVIEW: Papers submitted via the journal website will be checked by one of the editors: if these pass the first step, the papers will be sent to two reviewers for the peer-review process. After this, the author will be notified on the acceptance of his paper, or will be given suggestions on how to improve it.

RIGHTS: JLIS.it is published under a "Creative Commons Attribution License" (CC-BY).

Sommaro
Table of Contents
Vol. 3, n. 1 (Giugno/June 2012)

Saggi = Essays

Maria Violeta Bertolini	<i>The new ISBD Area 0 and the teaching of cataloging of non-book materials</i>	p. 4794-1
Emanuele Bellini, Chiara Cirinnà, Maurizio Lunghi ... [et al.]	<i>Il progetto National Bibliography Number Italia (NBN:IT). Un identificatore persistente a supporto del deposito legale nazionale delle risorse digitali</i>	p. 4789-1
Andrea Capaccioni, Giovanna Spina	<i>Italian HSS journals in Journal Citation Reports (JCR) and in SCImago Journal Rank (SJR): data and first analysis</i>	p. 4787-1
Paola Galimberti	<i>Qualità e quantità: stato dell'arte della valutazione della ricerca nelle scienze umane in Italia</i>	p. 5617-1
Erika Pasceri	<i>Analyzing rare diseases terms in biomedical terminologies</i>	p. 4783-1
Raffaele Guarasci	<i>Le preferenze di selezione verbali: un approccio computazionale</i>	p. 4786-1

Fare il punto = Making the Point on

Giuliano Genetasio	<i>The International Cataloguing Principles and their future</i>	p. 4936-1
--------------------	--	-----------

Dedicato a... = Dedicated to...

Carlo Revelli	<i>The catalogue (and the cataloguer): a defenseless entity?</i>	p. 5621-1
Alberto Cheti	<i>Dialogo sopra Il catalogo per soggetti di Carlo Revelli</i>	p. 4816-1
Pino Buizza	<i>Ri/leggere Revelli. Metonimia per una presentazione de Il catalogo per soggetti di Carlo Revelli</i>	p. 5616-1



The new ISBD Area 0 and the teaching of cataloging of non-book materials

Maria Violeta Bertolini

Introduction

The new International Standard Bibliographic Description (ISBD) Area 0 was published for the first time in the new ISBD Consolidated Edition in July 2011. With the purpose of introducing this new area, the General Material Designation (GMD) is taken as a precedent: a term or phrase that identified the material to catalogue, its addition was optional and had to be registered in brackets and after the title proper.

Different problems identified in the GMD took to the creation of the Area 0, among which have stood: “the steadily proliferation of the numbers and types of resources published in more than one physical and/or digital medium”, “the GMD terms reflected a confusing mix of physical format, class of material, form of carrier, and notation (e.g., Braille)”; [and], “the location of the GMD immediately following the title proper was seen as interrupting the logical order and sequencing of title information” (ISBD Review Group and IFLA Cataloguing Section. Standing Committee, “ISBD : International Standard Bibliographic Description: Area 0, Content Form and Media Type Area”).



For these and other reasons, in 2003 was created the Material Designations Study Group (MDSG) that proposed, between 2005 and 2007, the creation of "a separate, unique, high level component, a "content/carrier" or "content/medium" designation, mandatory for recording in bibliographic records" ("ISBD : International Standard Bibliographic Description: Area 0, Content Form and Media Type Area"). The result has been the incorporation of a new area, comprising two mandatory elements: Content form (with Content qualifications) and Media type.

The purpose of this new area is "to indicate at the very beginning of the description both the fundamental form or forms in which the content of a resource is expressed and the type or types of carrier used to convey that content so as to assist catalogue users in identifying and selecting resources appropriate to their needs" (*ISBD : International Standard Bibliographic Description (Consolidated ed.)*).

Brief history of GMD and its application through AACR2 and different versions of the ISBD

The need for GMD arises, between 1950 and 1960, when "non-book" or special materials started to be visible in library collections. First, it was identified the emergence of a document type designation for some materials in the *Rules for descriptive cataloging in the Library of Congress* published in 1952. These designations were incorporated into the Anglo-American Cataloguing Rules (AACR) when it was published in 1967 under the name of *qualifiers*, descriptive terms to identify the kind of material. In 1975, a review was made of certain chapters of AACR and material designations are renamed *medium designators*, and was enriched the vocabulary of terms that

identified bibliographic media and carriers (Chiverto). This review was undertaken by the American Library Association, the Canadian Library Association, and the Library of Congress (Tillett, "General Material Designations (GMDs) (4JSC/Chair/73)").

Moreover, "the term *general material designation* first appeared in 1977 ISBD (G) and 1977 ISBD (NBM) [and] the abbreviation 'GMD' appeared in the ISBD (G) examples" ("General Material Designations (GMDs) (4JSC/Chair/73)"). After the joint initiative of the Library of Congress and the Joint Advisory Committee on Non-book Materials, a list of designations was drawn up that could be equally useful to librarians and the general public and, in line with ISBD, the name of *medium designators* was changed to *general material designations* (Chiverto).

This list was formalized with the publication of AACR in 1978. In 1981, during a review of the ISBD, its location was set invariably next to the title proper, as an "early warning" about the type of resource of bibliographic description (Chiverto).

GMD was, according to ISBD Preliminary Consolidated Edition (ISBD Review Group and IFLA Cataloguing Section. Standing Committee, "ISBD : International Standard Bibliographic Description (Preliminary Consolidated ed.)"): "a term indicating broadly, the class of material to which a resource belongs". Meanwhile, the Anglo-American Cataloguing Rules, 2nd edition (AACR2) define it as "a term or phrase that expresses and identifies the material to be catalogued". Its addition was optional and should be registered in brackets, in the language of cataloging agency, singular, lowercase, and after the title proper.

The AACR2 (1.1C1) prescribed as optional the use of this element of description and for that made two lists available, corresponding to the designations used by British institutions on one hand, and on the other hand, those used by institutions in Australia, Canada, the

United States and Latin America:

- List 1 (for Britain): braille, sound recording, manuscript, cartographic material, graphic, microform, multimedia, music, object, motion picture, electronic resource, text and video-recording.
- List 2 (for North America, Australia and Latin America): braille, kit, chart, slide, technical drawing, diorama, picture, filmstrip, sound recording, game, toy, manuscript, cartographic material, microform, model, music, art original, motion picture, microscope slide, realia, art reproduction, electronic resource, activity card, flash card, text, transparency and videorecording.

In addition, the AACR2 indicated the possibility to add "(large print)" or "(tactile)" to any term of the List 2, for materials produced for the visually impaired. Also, "(braille)" could be added to any term of the List 2, other than braille or text.

Finally, the AACR2 (1.1C4) considered the possibility that a resource could contain parts that relate to materials covered by two or more categories of the selected list. In this case, the AACR2 indicated the designation "multimedia" or "kit" if none of these parts was the predominant element of the resource. In the preliminary consolidated edition of the ISBD (1.2), it was also stated that the GMD "is optional, but its inclusion is useful in integrated catalogues or bibliographies".

There are very interesting articles on the emergence and evolution of the GMD over the past 60 years. Particularly, Tillett ("General Material Designations (GMDs) (4JSC/Chair/73)"), Guerrini ("GMD: its Function and its History") and Salamanca Chiverto ("La Categorización de Contenidos y Medios en la Descripción Bibliográfica: la Designación General de Material (DGM) y su Pasado, Presente y

Futuro en la Era Digital”), perform a detailed analysis of the emergence of GMD and they all conclude in the need to be modified or rethought.

According to Guerrini (“GMD: its Function and its History”), “the use of GMD must be considered difficult for the following reasons:

1. its hybrid, “classificatory” and functional nature;
2. the existence of many different lists;
3. the incomprehensibility of some terms for cataloguers and/or users;
4. the difficult application for multiple characteristic resources, because many terms are too general;
5. its optional application;
6. its display in the OPAC”.

For hybrid nature, Guerrini (“GMD: its Function and its History”) highlights Delsey’s observations about “the large inconsistency of the list of GMDs, which includes terms referring to the class of material, but also to physical format, form of carrier and notation (e.g. braille)”. It also suggests that “issues 2, 3 and 5 could easily be solved, if only an answer could be found to the question of the real nature of GMD and to its consistent, consequent application”.

Similarly, Salamanca Chiverto (“La Categorización de Contenidos y Medios en la Descripción Bibliográfica: la Designación General de Material (DGM) y su Pasado, Presente y Futuro en la Era Digital”) notes “the prevailing terminological confusion surrounding the GMD, manifested in the formation of lists of terms prepared according to heterogeneous classification criteria, and in which there is coexistence of different linguistic levels alternating terms of artificial

nature with other expressed in a common vocabulary and accessible for the average user”.

For her part, Tillett (“General Material Designations (GMDs) (4JSC/Chair/73)”) notes after a careful investigation on GMDs, that “there is no ‘cataloging theory’ behind their use. They were initially created at a time when public libraries were beginning to acquire and catalog nonprint materials. These institutions wanted the catalog records for the nonprint materials to be included in their general card catalogs and they also wanted these catalog records to be easily identified as ‘nonprint’”.

Creation of the new Area 0

In 2003, at the first IFLA Meetings of Experts on an International Cataloguing Code (IME-ICC) in Frankfurt, Huthwaite (“Class of Materials Concept and GMDs (4JSC/ACOC rep/1)”) presented a document that analyzed the problems identified in the class of material concept and GMD. In this document, Huthwaite refers to two suggestions for fundamental change introduced by Tillett (“General Material Designations (GMDs) (4JSC/Chair/73)”) in her investigation of the history of the GMDs: “firstly, use a device in the bibliographic record representing the mode of expression; and secondly, moving terms representing physical format and form of carrier to area 5 (physical description) or to notes”.

The Joint Steering Committee for Revision of AACR2, on May 2002, had discussed the possibility of implementing Tillett’s suggestions and indeed, “terms from List 2 in AACR2 at the expression level were identified [and] the most of the remaining terms could be relocated to area 5, some of which could be used as qualifiers (e.g., “braille” and “electronic resource”)” (Huthwaite). In this sense, Huthwaite states that “when terms and concepts from Functional

Requirements for Bibliographic Records (FRBR) are introduced into AACR2, it will be logical and consistent to consider the general material designation as an expression-level indicator, and the specific material designation as a manifestation-level indicator". She even mentions the possibility to "incorporate the expression-level indicator in the textual part of the bibliographic record, possibly in a new area". Finally, Huthwaite recommends the creation of "a group that compiles a complete list of expression-level indicators representing all modes of expression, including both those relocated from the list of GMDs and any others identified in FRBR". Also, recommends that "a group be formed to make recommendations on a rationalization of the terms used in area 5, and on how statements could be constructed that allow for the description of all aspects of an item/resource at the manifestation level". This group, she adds, "could be the same group examining the deconstruction of GMDs".

In 2003, the ISBD Review Group created the MDSG to investigate the Specific Material Designations (SMD) and the GMD applied to multiple formats and media. This group emerged in response to concerns that had been raised regarding the following issues:

- the steadily proliferation of the numbers and types of resources published in more than one physical and/or digital medium;
- the GMD terms reflected a confusing mix of physical format, class of material, form of carrier, and notation (e.g., Braille),
- the location of the GMD immediately following the title proper was seen as interrupting the logical order and sequencing of title information (ISBD Review Group and IFLA Cataloguing Section. Standing Committee, "ISBD : International Standard Bibliographic Description: Area 0, Content Form and Media Type Area").

On this last point, Salamanca Chiverto (“La Categorización de Contenidos y Medios en la Descripción Bibliográfica: la Designación General de Material (DGM) y su Pasado, Presente y Futuro en la Era Digital”) observes that “regarding the functional aspects and location in the bibliographic record, we see also a contradiction: an item designed for the selection of library resources through classification of its formal nature, has been incorporated into the record as “intruder” in a descriptive field, the title and statement of responsibility area”. Also, the MDSG suggested the review of the situation of the GMD as an optional element within the ISBD, taking into account its relative importance to the catalog user in identifying and selecting appropriate resources.

Escolano Rodríguez (“Edición Preliminar de la ISBD Consolidada”) reports in detail the process of study and exchange that took place between 2005 and 2007 to reach the definition of Area 0, and emphasizes that the MDSG worked on the basis of the agreement on the importance and primacy that GMD had as an “early warning device” for catalog users. At the ISBD Review Group IFLA 2005 meetings in Oslo, the MDSG proposed the creation of “a separate, unique, high level component, a “content/carrier” or “content/medium” designation, mandatory for recording in bibliographic records” (“ISBD : International Standard Bibliographic Description: Area 0, Content Form and Media Type Area”). Some of the following assumptions were incorporated in the proposal:

- should, likewise, serve as an “early warning” or filtering device, assisting catalogue users;
- the structure of terms should be logical, the categories mutually exclusive, and the terminology clear, unambiguous, and readily understood by users;
- existing ISBD terminology should be retained wherever pos-

sible and applicable, to support compatibility with legacy records, and to minimize the number of new terms requiring extensive systems updates;

- carrier categories in particular should be sufficiently flexible to accommodate new formats and types of resources as they evolve or are identified;
- terminology should be as compatible as possible or appropriate with that used by other metadata communities (e.g., publishing, museums, archives) to support interoperability.

The ISBD Review Group commissioned the MDSG to prepare a final text. In 2007 IFLA meetings in Durban, the Preliminary Consolidated Edition of the ISBD had been published, and the MDSG had drafted the proposal for a content/carrier component for ISBD Review Group discussion.

In November 2008 was distributed for review at global level, the draft proposal for a new ISBD area for description, with the number 0 (zero) and entitled "Content Form and Media Type Area". As states Escolano Rodriguez ("Estado de la Cuestión de ISBD"), the decision to designate with the number 0 (zero), the content form and media type area, was the intention of "not to alter the previous numerical structure firmly rooted, established in numbered areas, which numeric order reflected the relative importance of information encompassing the areas". Furthermore, "according to MDSG recommendation: the information contained in this area occupies a prominent place as a first call user's attention, to serve for an initial selection of resources". After an analysis of the responses received and further consideration by the ISBD Review Group, by 2009 the final text was approved and referred to the Standing Committee of the IFLA Cataloguing Section for publication in a preliminary edition independently of ISBD Consolidated Edition. By 2009, the

preliminary version of Area 0 was published, composed of three mandatory elements, if they are applicable: Content form, Content qualification, and Media type. The purpose of this new area is "to indicate at the very beginning of the description both the fundamental form or forms in which the content of a resource is expressed and the type or types of carrier used to convey that content so as to assist catalogue users in identifying and selecting resources appropriate to their needs" (*ISBD : International Standard Bibliographic Description (Consolidated ed.)*). In May 2010 the draft of the definitive edition of the ISBD Consolidated was published for worldwide review until July 11, 2010. Among other general amendments to the text, this draft no longer had the GMD in Area 1 and included the Area 0. Gordon Dunsire published later on versions in English and Spanish of the vocabulary of ISBD Area 0 Content forms, Content qualifications and Media types in the Open Metadata Registry. Finally, in July 2011, the publication of the ISBD consolidated edition was released, edited by the Standing Committee of the IFLA Cataloguing Section. To date it is only available in English language, in printed and electronic formats, published by De Gruyter Saur.

Structure of Area 0

The introductory note of the ISBD Consolidated (*ISBD : International Standard Bibliographic Description (Consolidated ed.)*) Area 0 stands: "the increasing variety and complexity of resource content and types of media, along with a growing diversity of users with particular information needs, require greater specificity in describing content form and media type early in the record". A major difference between the version of Area 0 independently published in 2009 and the draft of the consolidated edition published in 2010 is that, in the final version, Area 0 had been reduced to two elements

(instead of three): Content form and Media type. The qualifiers were included in the first element, rather than appearing as a separate item. Furthermore, the most significant difference with regard to the former GMD is the mandatory nature of both elements. In the preliminary edition of Area 0 the media type was mandatory only if it was applicable, meaning it was assumed that the resource was accessible "unmediated" if it required no mediating device to use or perceive the resource and therefore not media type was recorded. Currently, the media type is always required. For example, in the case of printed books (content form "Text") media type "unmediated" should always be included → Text (visual or tactile) : unmediated.

Elements of Area 0: **1. Content form:** one or more terms reflecting the fundamental form or forms in which the content of a resource is expressed, to which may be added one or more content qualifications, specifying the type, sensory nature, dimensionality, and/or presence or absence of motion for the resource being described. **2. Media type:** indicating the type or types of carrier used to convey the content of the resource.

Prescribed punctuation:

Content form (content qualification) : media type

Different content qualification terms are applicable to one content form: Content form (content qualification ; content qualification) : media type.

Different content forms are contained in one media type: Content form. Content form (content qualification) : media type *or* Content form (content qualification). Content form (content qualification) : media type.

A resource consists of different media types and contains different content form : Content form (content qualification) : media type
+ Content form (content qualification) : media type

Prescribed source: The resource itself.

Content form (0.1):¹ Content form categories reflect the fundamental form or forms in which the content of a resource is expressed. It includes one or more terms from a predefined list in the language and script chosen by the cataloguing agency.

For resources of mixed content where there is no predominant part of the resource (i.e. all parts are equally prominent or important), all terms that are applicable to the resource being described are recorded, in alphabetical order. Exceptionally, when three or more forms are applicable, the term multiple content forms may be given.

Dataset → content expressed by digitally-encoded data intended to be processed by a computer.

Examples: numeric data, environmental data etc., used by applications software to calculate averages, correlations etc., or to produce models etc., but not normally displayed in its raw form.

Excluded are digitally recorded music [see music], language [see spoken word], sounds [see sounds], computer-reproduced images [see image] and text [see text].

Cartographic database → *Dataset (cartographic)* : *electronic*

Database of trade statistics → *Dataset* : *electronic*

¹ISBD Content Form: <http://metadataregistry.org/vocabulary/show/id/113.html>.

Image → content expressed through line, shape, shading etc.; an image can be still or moving, in two or three dimensions.

Examples: art reproductions, maps, raised relief maps, photographs, remote-sensing images, stereographs, motion pictures, and lithographs.

Ballet (photographs and noted choreography)→ Image (still ; 2-dimensional ; visual). Movement (notated ; visual) : unmediated

Drawing with plush→ Image (still ; 2-dimensional ; visual ; tactile) : unmediated

GPS (Global Positioning System)→ Image (cartographic ; moving ; 2-dimensional ; visual). Spoken word : electronic

Picture→Image (still ; 2-dimensional ; visual) : unmediated

Map in slides→Image (cartographic ; still ; 2-dimensional ; visual) : projected

Map on website with geographic information→Image (cartographic). Text (visual) : electronic

Maps on video→Image (cartographic ; moving ; 2-dimensional) : video

Movie (film reels)→Image (moving ; 2-dimensional) : projected

Poster→Image (still ; 2-dimensional ; visual) : unmediated

Printed map→Image (cartographic ; still ; 2-dimensional ; visual) : unmediated

Relief printed map→Image (cartographic ; still ; 2-dimensional ; visual ; tactile) : unmediated

Stereography→Image (still ; 3-dimensional ; visual) : stereographic

Sticker with plush and smell→Image (still ; 2-dimensional ; visual ; tactile ; olfactory) : unmediated

Video recording (3D Blu-ray)→Image (moving ; 3-dimensional) : video

Video recording (DVD)→Image (moving ; 2-dimensional) : video

Visual projection (slides)→Image (still ; 2-dimensional ; visual) : projected

Movement → content expressed through motion, i.e. the act or process of changing the position of an object or person.

Examples: dance notation, stage actions, or choreography.

Excluded are moving images, such as motion pictures [see image]

Braille notated choreograph → *Movement (notated ; tactile) : unmediated*

Printed notated choreography → *Movement (notated ; visual) : unmediated*

Music → content expressed through ordered tones or sounds in succession, in combination, and in temporal relationships to produce a composition.

Music can be written (notation), performed, or recorded in analogue or digital formats as vocal, instrumental, or mechanical sounds having rhythm, melody, or harmony.

Examples: : written music, such as scores or parts, and recorded music, such as concert performances, opera, and studio recordings.

Braille printed score → *Music (notated ; tactile) : unmediated*

Music (vinyl record) → *Music (performed) : audio*

Music on a local disc → *Music (performed) : electronic*

Music on an mp3 player → *Music (performed) : audio*

Music on CD-ROM → *Music (performed) : audio*

Printed score → *Music (notated ; visual) : unmediated*

Score in PDF → *Music (notated ; visual) : electronic*

Object → content expressed through either naturally occurring entities or human-made or machine-manufactured artifacts; also referred to as three-dimensional structures or realia.

Examples of artifacts: sculpture, models, games, coins, toys, buildings, equipment, clothing, cultural, and other items.

Examples of natural occurring entities: fossils, rocks, insects, biological specimens on slides etc.

Examples of cartographic objects: globes, relief models, and cross-sections intended to be viewed from the side in three-dimensional form.

Raised relief maps are excluded [see image].

Sculpture → *Object (visual)* : *unmediated*

Cartographic resource (relief globe) → *Object (cartographic ; visual ; tactile)* : *unmediated*

Cartographic resource (globe) → *Object (cartographic ; visual)* : *unmediated*

Program → content expressed through digitally-encoded instructions intended to be processed and performed by a computer.

Examples: operating systems, application software etc.

Microsoft Office 2010 Software (CD-ROM) → *Program* : *electronic*

Sounds → content expressed through sounds made by animals, birds, naturally occurring sources of noise, or such sounds simulated by the human voice or digital (or analogue) media.

Examples: recordings of bird songs, animal calls, and sound effects.

Excluded are recorded music [see music] and recorded human speech [see spoken word].

Sound of the whales (Cassette)→Sounds : audio

Spoken word → content expressed through the sound of the human voice talking.

Examples: talking books, radio broadcasts, oral history recordings, and audio recordings of plays, whether recorded in analogue or digital form.

Audio book (CD-ROM) →Spoken word : audio

Text → content expressed through written words, symbols and numbers.

Examples: books (printed or electronic), correspondence, databases of journals and microfilmed newspapers.

Braille printed book→Text (tactile) : unmediated
Continuing resource on CD-ROM→Text (visual) : electronic
Continuing resource on microform→Text (visual) : microform
Continuing resource printed and on microform→Text (visual) : unmediated + Text (visual) : microform
E-book→Text (visual) : electronic
Power Point presentation→Text (visual) : electronic
Printed book→Text (visual) : unmediated
Printed continuing resource→Text (visual) : unmediated
Slides presentation→Text (visual) : projected
Travel Guide on CD-ROM→Text (visual). Image (cartographic) ;

visual) : *electronic*

Website→*Text (visual)* : *electronic*

Multiple content forms → mixed content where three or more forms are applicable. Alternatively, all terms that are applicable to the resource being described can be recorded in alphabetical order.

Multimedia → *Multiple content forms* : *multiple media*

Multimedia kit of "Microsoft Windows 98" composed of: installation guide on VHS, CD-ROM software installation and user manual

→ *Image (movement ; 2-dimensional ; visual)* : *video* + *Program* : *electronic* + *Text (visual)* : *unmediated*

Other content form → content expressed through a different content form of: database, image, movement, music, object, program, sound, spoken word, text or multiple content forms.

Content qualification (0.1.1): A content form category is expanded by one or more content qualification sub-categories that are applicable to the resource being described. Content qualifications specify the type, presence or absence of motion, dimensionality, and sensory nature for the resource being described. Terms are taken from predefined lists, or their equivalents in the language and script of the cataloguing agency, using as many terms as are appropriate to expand on or further clarify the corresponding content form category.

Qualifications are added if different options are possible for a given content form.

Examples: Text (visual) o Text (tactile); Image (still ; 2-dimensional) o Image (still ; 3-dimensional); Image (movement ; 2-dimensional) o Image (movement ; 3-dimensional).

Specification of type (0.1.1.2)²

- Cartographic → content representing the whole or part of the Earth or any celestial body at any scale. It includes maps, atlases, globes, relief models etc.
- Notated → content expressed through a notational system for artistic purposes (e.g. music, dance, and staging) intended to be perceived visually.
- Performed → content expressed in an audible or visible form, at a given time, and recorded in a resource. It includes recorded performances of music or movement, computer-generated music etc.

For use with Content form “image” only:

Specification of motion (0.1.1.3)³: The content form image is qualified to show the perceived presence or absence of movement in the image content of a resource.

- Movement → Image content that is perceived to be in motion, usually by means of a quick succession of images.
- Still → Image content that is perceived to be static.

Specification of dimensionality (0.1.1.4)⁴: The content form image is qualified to show the number of spatial dimensions in which the image content of a resource is intended to be perceived.

- 2-dimensional → Image content that is intended to be perceived in two dimensions.

²ISBD Content Qualification of Type: <http://metadataregistry.org/vocabulary/show/id/115.html>

³ISBD Content Qualification of Motion: <http://metadataregistry.org/vocabulary/show/id/116.html>

⁴ISBD Content Qualification of Dimensionality: <http://metadataregistry.org/vocabulary/show/id/117.html>

- 3-dimensional → Image content that is intended to be perceived in three dimensions.

Sensory specification (0.1.1.5):⁵ This attribute indicates a human sense through which the content of a resource as published is intended to be perceived.

- Aural → Content that is intended to be perceived through hearing.
- Gustatory → Content that is intended to be perceived through taste.
- Olfactory → Content that is intended to be perceived through smell.
- Tactile → Content that is intended to be perceived through touch.
- Visual → Content that is intended to be perceived through sight.

Media type (0.2):⁶ media type categories record the type or types of carrier used to convey the content of the resource. Categories generally reflect the format of the storage medium and housing of a carrier in combination with the type of intermediation device required to render, view, run etc., the content of a resource.

Unmediated → *For media for which no mediation device is necessary.*

Audio → *For audio player-enabled resources.* Media used to store recorded sound, designed for use with a playback device such as a turntable, audiocassette player, CD player, MP3 player, or iPod;

⁵ISBD Content Qualification of Sensory Specification: <http://metadataregistry.org/vocabulary/show/id/118.html>

⁶ISBD Media Type: <http://metadataregistry.org/vocabulary/show/id/114.html>

includes media used to store digitally encoded as well as analogue sound.

Electronic → *For computer-enabled resources.* Media used to store electronic files, designed for use with a computer, includes media that are accessed remotely through file servers, as well as direct-access media such as computer tapes and disks.

Microform → *For microform reader-enabled resources.* Media used to store reduced-size images, not readable to the human eye, and designed for use with a device such as microfilm or microfiche reader; includes both transparent and opaque micrographic media.

Microscopic → *For microscope-enabled resources.* Media used to store minute objects, designed for use with a device such as a microscope to reveal details invisible to the naked eye.

Projected → *For projector-enabled resources.* Media used to store moving or still images, designed for use with a projection device such as a motion picture projector, slide projector, or overhead projector, includes media designed to project both two-dimensional and three-dimensional images.

Stereographic → *For stereograph viewer-enabled resources.* Media used to store pairs of still images, designed for use with a device such as a stereoscope or stereograph viewer to give the effect of three dimensions.

Video → *For video player-enabled resources.* Media used to store moving or still images, designed for use with a playback device such as a videocassette player or DVD player, includes media used to store digitally encoded as well as analogue images.

Multiple media → *For resources composed of mixed media* where three or more media types are applicable (used exceptionally).

Other media → *If none of the other terms listed applies* to the type of carrier and intermediation device required to convey, use, or perceive the content of the resource being described.

For resources of mixed media where there is no predominant part of the resource (i.e. all parts are equally prominent or important) → all terms that are applicable to the resource are recorded.

For works of mixed media where one part of the resource is predominant and other media types are minimal or incidental → media types that are not predominant may be omitted.

Compatibility with formats, standards and conceptual models: MARC, Resource Description & Access (RDA) and FRBR

“In other words, the problem is: how to manage the relationship between FRBR, the new international cataloging principles (ICP), the consolidated edition of ISBDs and cataloging codes (e.g., RDA, Resource Description and Access)? The bibliographic universe can be efficiently organized with the sole condition of a constant interaction between theory and practice.

Nowadays, we attend a strong influence and pressure of a theoretical model on all established practices, to the point that they seem totally inappropriate for the (new?) information needs. This influence suggests breaking with past practice and making room for entirely new models and tools. The transition from past to future, however, doesn't prevent us from thinking of the present. Therefore, it is urgently needed the search for a solution to the problem of the management and the interrelationships between FRBR, ISBD and RDA.” (Bianchini and Guerrini, “Quis custodiet ipsos custodiet?: observaciones sobre la relación entre FRBR, ICP, ISBD y RDA”)

Compatibility and interoperability between different standards, conceptual models and formats of bibliographic description is essential in relation to compliance with international principles of cataloging and universal bibliographic control. As it stands Guerrini, we are

in a time of change and transitions given not only in the cataloging world but also in the world of information technology and publishing. In this sense, it is important to ensure compatibility, without this meaning to stop taking risks. It is not always successful the phrase *if it isn't broke don't fix it*. In some cases, even if everything is working, it might work better. However, it is important to work together, listen to opinions and perform tests before making rushed decisions, especially given the costs and difficulties that present some realities of many libraries around the world. At this point, we find four groups of key elements to consider: ISBD, RDA/AACR, FRBR/FRAD/FRSAD and MARC. Although all of these elements are closely related in its origin, the great challenge for the working groups at international level in their development is to achieve greater compatibility. The aim is that all these elements constitute and complement as a cataloger's "toolkit" and therefore, we cannot speak nowadays of a particular element of the ISBD (as Area 0) without asking: where is this located in MARC, what establishes RDA about it?, and, with which FRBR entities and attributes corresponds?

Next, we will attempt to answer these questions.

Inclusion of Area 0 in MARC format

The inclusion of Area 0 and its vocabularies still does not have a final location in MARC21 Bibliographic format. At first, would be applicable the fields created for the consignment of the RDA vocabularies for content, media and carrier types. The fields, created in 2009 for these three items, were included in the section "Physical Description etc. Fields (3XX)" ⁷ MARC21 located in the numbers:

⁷MARC 21 3XX - Physical Description etc. Fields - General Information:<http://www.loc.gov/marc/bibliographic/bd3xx.html>

336 (Content Type), 337 (Media Type) and 338 (Carrier Type).⁸

In contrast, in the case of UNIMARC, as explained by Le Pape (“FRBR en UNIMARC: Propositions du Cfu Validées par le PUC”) the decision was to include Area 0 through two coded fields and a text field. First, the coded data fields: 181 field (Content Characteristics of Expression) for ISBD Content form codes and ISBD Content qualification codes, and, the 182 field (Media Type) for ISBD Media type codes. Furthermore, in order to maintain a text field that respects natural structure of Area 0 in its entirety, field 203 was created (Content form and Media type). This field includes 3 subfields, each one with a portion of Area 0: \$ a Content form; \$ b Content qualification; \$ c Media type. This field is optional when fields 181 and 182 are present, due to having redundant data, although in another format.

MARC 21 fields for content type, media type and carrier type and sources of terms and codes RDA

To accommodate ISBD Area 0, but using other terms, MARC21 has two ways in which to express Content form and Media type.

On one hand, the Leader Field /06 and the Control Fields 007 and 008 that are applicable to all types of documents. These are related to the ISBD Area 0, with the 3XX fields and some of 5XX note fields, since they define in encoded form many elements about content, media and carrier types.

- The Leader, a fixed field that comprises the first 24 character positions (00-23) of each bibliographic record and consists of data elements that contain numbers or coded values that define the parameters for the processing of the record. In

⁸RDA in MARC (September 2011): <http://www.loc.gov/marc/RDAinMARC29-9-12-11.html>

particular, the Leader Field /06⁹ (Type of record) is a one-character alphabetic code used to define the characteristics and components of the record.

- Control field 007¹⁰ (Physical Description Fixed Field-General Information) contains special information about the physical characteristics in a coded form. The information may represent the whole item or parts of an item such as accompanying material.
- Control field 008¹¹ (Fixed-Length Data Elements-General Information) contains forty character positions (00-39) that provide coded information about the record as a whole and about special bibliographic aspects of the item being cataloged. These coded data elements are potentially useful for retrieval and data management purposes.

On the other hand, we can find fields: 336 (Content Type), 337 (Media Type) and 338 (Carrier type), mentioned above. They are described below:

Field 336 - Content type:¹² reflects the form of communication through which a work is expressed. Used in conjunction with Leader /06 (Type of record), which indicates the general type of content of the resource.

Field 337 - Media type:¹³ reflects the format of the storage medium and housing of a carrier in combination with the media type (which

⁹<http://www.loc.gov/marc/bibliographic/bdleader.html>

¹⁰<http://www.loc.gov/marc/bibliographic/bd007.html>

¹¹<http://www.loc.gov/marc/bibliographic/bd008.html>

¹²MARC 21 336 - Content Type (R) (October 2009): <http://www.loc.gov/marc/bibliographic/bd336.html>

¹³MARC 21 337 - Media Type (R) (October 2009): <http://www.loc.gov/marc/bibliographic/bd337.html>

indicates the intermediation device required to view, play, run etc., the content of a resource). Used as an alternative to or in addition to the coded expression of media type in Control Field 007/00 (Category of material).

Field 338 - Carrier type:¹⁴ reflects the format of the storage medium and housing of a carrier in combination with the media type (which indicates the intermediation device required to view, play, run etc., the content of a resource). Used as an alternative to or in addition to the coded expression of carrier type in field 007/01 (Specific material designation).

Subfield codes for field 336:

\$a –Content type term: content type of the work being described.

336 ##\$aperformed music\$bprm\$2rdacontent

336 ##\$atwo-dimensional moving image\$btdi\$2rdacontent

\$b –Content type code: Code representing the content type of the work being described.

336 ##\$aperformed music\$bprm\$2rdacontent

336 ##\$atext\$btxt\$2rdacontent\$3liner notes

Subfield codes for field 337:

\$a –Media type term: Term for the category of media used to convey the content of the resource.

337 ##\$aaudio\$bs\$2rdamedia

337 ##\$avideo\$bv\$2rdamedia

\$b –Media type code: Code for the category of media used to convey the content of the resource.

337 ##\$aaudio\$bs\$2rdamedia

337 ##\$aunmediated\$bn\$2rdamedia\$3liner notes

¹⁴MARC 21 338 - Carrier Type (R) (October 2009): <http://www.loc.gov/marc/bibliographic/bd338.html>

Subfield codes for field 338:

\$a –Carrier type term: Term for the category of carrier used to convey the content of the resource.

338 ##\$aaudio disc\$bsd\$2rdacarrier

338 ##\$avideodisc\$bvd\$2rdacarrier

\$b –Carrier type code: Code for the category of carrier used to convey the content of the resource.

338 ##\$aaudio disc\$bsd\$2rdacarrier

338 ##\$avideodisc\$bvd\$2rdacarrier

Subfields present in the three fields:

\$2 –Source: MARC code that identifies the source of the term or code used to record the content/media/carrier type information (Genre/Form Code and Term Source Codes).

336 ##\$athree-dimensional moving image\$btdm\$2rdacontent

337 ##\$avideo\$bv\$2rdamedia

338 ##\$asheet\$bnb\$2rdacarrier\$3liner notes

\$3 –Materials specified: Part of the described materials to which the field applies.

336 ##\$atext\$2rdacontent\$3liner notes

337 ##\$aunmediated\$2rdamedia\$3liner notes

338 ##\$asheet\$2rdacarrier\$3liner notes

Multiple content types, media types and carrier types from the same source vocabulary or code list may be recorded in the same field (336, 337 or 338, accordingly) in separate occurrences of subfield \$a (term) and subfield \$b (code). Terms from different source vocabularies are recorded in separate occurrences of the field.

110 2# \$a System of a Down (Musical group)	110 2# \$a System of a Down (Musical group)
245 10 \$a Hypnotize / \$c System of a Down	245 10 \$a Hypnotize /\$c System of a Down
336 ## \$a performed music \$2 rdacontent	336 ## \$b prm \$2 rdacontent
336 ## \$a two-dimensional moving image \$2 rdacontent	336 ## \$b tdi \$2 rdacontent
336 ## \$a text \$2 rdacontent \$3 liner notes	336 ## \$b txt \$2 rdacontent \$3 liner notes
337 ## \$a audio \$2 rdamedia	337 ## \$b s \$2 rdamedia
337 ## \$a video \$2 rdamedia	337 ## \$b v \$2 rdamedia
337 ## \$a unmediated \$2 rdamedia \$3 liner notes	337 ## \$b n \$2 rdamedia \$3 liner notes
338 ## \$a audio disc \$2 rdacarrier	338 ## \$b sd \$2 rdacarrier
338 ## \$a video disc \$2 rdacarrier	338 ## \$b vd \$2 rdacarrier
338 ## \$a sheet \$2 rdacarrier \$3 liner notes	338 ## \$b nb \$2 rdacarrier \$3 liner notes

Table 1: Example of a portion of a bibliographic record representing the manifestation using RDA terms or codes (MARC 21 Bibliographic/Authority 336 - CONTENT TYPE (R) (Draft of approved field 2009): <http://www.loc.gov/marc/changes-rda-336.html>; MARC 21 Bibliographic/Authority 337 - MEDIA TYPE (R) (Draft of approved field 2009): <http://www.loc.gov/marc/changes-rda-337.html>; MARC 21 Bibliographic/Authority 338 - CARRIER TYPE (R) (Draft of approved field 2009): <http://www.loc.gov/marc/changes-rda-338.html>)

Term and code list for RDA content types,¹⁵ media types¹⁶ and carrier types¹⁷

To complete these fields (336, 337 and 338), is required the use of controlled lists of terms and/or codes designated to record information about content, media and carrier types.

There are several sources of lists of terms and codes, each code consists of a maximum of 12 lowercase alphabetic characters, that will be placed in the subfield \$ 2 of fields 336, 337 and 338 (MARC21

¹⁵Term and code list for RDA content types: rdacontent <http://www.loc.gov/standards/valuelist/rdacontent.html>

¹⁶Term and code list for RDA media types: rdamedia: <http://www.loc.gov/standards/valuelist/rdamedia.html>

¹⁷Term and code list for RDA carrier types: rdacarrier: <http://www.loc.gov/standards/valuelist/rdacarrier.html>. More details in the appendix, available at <http://leo.cilea.it/index.php/jlis/article/downloadSuppFile/4794/181>.

Bibliographic) in the case of being used.

Genre/Form Sources contains a list of sources of terms and/or codes indicating the genre, form, and/or physical characteristics of the materials being described in bibliographic records and assigns a code to each source.

The lists relating to Area 0 are:

- RDA content types that indicate the form of communication through which a work is expressed. Examples include: cartographic dataset, notated music, or text.
- RDA media types that specify the general type of intermediation device required to view, play, run etc., the content of a resource audio, stereographic, or video.
- RDA carrier types that specify the format of the storage medium and housing of a resource. Examples include audiocassette, microfiche, or online resource.

The three lists contain RDA terms and corresponding codes: three-character lower case alphabetic code for content type, one-character lower case alphabetic code for media type and two-character lower case alphabetic code for carrier type. These types can be recorded in MARC Bibliographic and Holdings records in textual and/or coded form. In MARC21 Bibliographic records, content, media and carrier types may also be recorded, sometimes in a more granular and sometimes in a more general form, with one-character codes in MARC Bibliographic Leader/06 (Type of material), and in Control Fields 007/01 (Specific material designation) and 007/00 (Category of material).

Mapping ISBD Area 0 vocabularies to RDA content, media and carrier vocabularies via the RDA/ONIX Framework (ROF) for resource categorization

The ISBD Review Group has suggested that the alignment between the vocabularies of the ISBD and RDA should be done within the ROF, “a scheme that identifies and defines two distinct sets of attributes: one for the intellectual or artistic content of an information resource, and the other for the means and methods by which such content is carried (...) for supporting the needs of libraries and the publishing industry for categorizing resources by their content and carrier, and of facilitating interoperability between the metadata produced by those communities” (Dunsire, “Distinguishing Content from Carrier: The RDA/ONIX Framework for Resource Categorization”).

The ROF is an initiative that started in 2005 between the Joint Steering Committee (JSC) for Revision of AACR and representatives of the publishing industry in the UK, and was followed up by the JSC for Development of RDA and organizations responsible for developing Online Information Exchange (ONIX), with support from the British Library. The ONIX is “a standard for the use of publishers in distributing digital metadata about their products”. It includes “a standard set of codes to be used in the metadata, including elements describing product content and carrier” (“Distinguishing Content from Carrier: The RDA/ONIX Framework for Resource Categorization”). The objective of ROF initiative is “to develop a framework for categorizing resources in all media that will support the needs of both libraries and the publishing industry and will facilitate the transfer and use of resource description data across the two communities” (“RDA/ONIX Framework for Resource Categorization”). At the second meeting of the ISBD/XML Study Group held in August 2010 in Gothenburg (Sweden), Dunsire (“Analysis of

Content and Carrier Designators in the ISBD Consolidated Edition with Respect to the RDA-ONIX Framework”) presented an analysis of content and carrier designators in the ISBD consolidated edition with respect to the RDA/ONIX Framework. The analysis was a mapping between elements of the ISBD Area 0 (using as a base at the time the draft consolidated edition 2010), with the RDA designations for these elements and the corresponding category in ROF. For media types, the ROF designation, the RDA vocabularies and the terminology used in the ISBD Area 0 are very compatible. Table 2 is prepared by Dunsire (“Analysis of Content and Carrier Designators in the ISBD Consolidated Edition with Respect to the RDA-ONIX Framework”), detailing the partial mapping:

ISBD label	RDA label	ROF label
audio	audio	audio player
electronic	computer	computer
microform	microform	microform reader
microscopic	microscopic	microscope
multiple media	-	-
other media	-	-
projected	projected	projector
stereographic	stereoscopic	stereoscope
video	video	audiovisual player
unmediated	unmediated	not required

Table 2

Dunsire states: “the ISBD media type is equivalent to the RDA media type. Neither can be mapped to a base carrier category in ROF, which contains values for the attributes StorageMediumFormat, HousingFormat, and IntermediationTool. Instead, ISBD and RDA media types map to the single attribute IntermediationTool” (“Analysis of Content and Carrier Designators in the ISBD Consolidated Edition with Respect to the RDA-ONIX Framework”).

In the case of area 0, this is because the ISBD has area 5 "Material Description Area" that "relates a resource to a statement including the specific material designation and extent, other physical details, the dimensions, and the accompanying material statement",¹⁸ to describe aspects of the manifestation and carrier. Therefore, it is not expected a mapping between RDA type of carrier vocabulary and ISBD area 0. In the case of area 5, it has been preferred to allow each cataloging agency the use of appropriate vocabulary for their users, so there has been no normalization of specific material designations (DEM). In his analysis, Gordon Dunsire ("Analysis of Content and Carrier Designators in the ISBD Consolidated Edition with Respect to the RDA-ONIX Framework") makes observations and recommendations involving amendments to the ISBD. Some of these included:

1. expand the ISBD 0.1.1 instruction to alert users to potential ambiguity if qualifiers are not added because they are assumed to be implicit in the content form term;
2. redraft the definitions of "image" and "object" to ensure that there is no overlap;
3. add an explanation to ISBD area 0 to show how content and carrier designators can be used in practice in an online environment to meet the needs of users and use examples in ISBD area 0 that are not ambiguous and clearly support the utility of area 0, or which illustrate problems when implicit assumptions are made.
4. the ISBD Review Group should monitor use of the media type "other media" to inform the addition of new values to the ROF IntermediationTool attribute and ISBD media types;

¹⁸Element Set: ISBD elements: <http://iflstandards.info/ns/isbd/elements.rdf>

5. redraft the instruction concerning mixed content at ISBD 0.1 to clarify the treatment of mixed content and media resources;
6. allow the recording of any applicable content and media designators, rather than just predominate ones. Although ROF does not require all applicable content and carrier categories to be recorded for a resource, it does provide a set of five values to indicate the extent of applicability of a category to a resource (full, predominate, substantial, some, none). ISBD does not provide explicit indicators. The absence of a categorization of this kind for ISBD creates confusion about the ambiguity of the word "predominant" to determine whether to include or not a particular Content form.

Some of these problems were already observed and resolved with the publication of the new ISBD Consolidated in 2011, mainly due to substantial changes in the mandatory nature of the two components of Area 0 (Content form and media type). For example in the case of media type "unmediated" instead of being implied by the absence of another media type, its inclusion is now mandatory. Also, in the new edition, the content qualifications are mandatory for content forms that have different options, allowing further specification. It includes also more specifically detailed instructions for content forms and qualifiers, and the "Image" and "Object" definitions were reviewed to avoid ambiguity. Regarding the use of "multiple content forms" the ISBD Consolidated (ISBD Review Group and IFLA Cataloguing Section. Standing Committee, *ISBD : International Standard Bibliographic Description (Consolidated ed.)*) provides its exceptional use, if absolutely necessary, and recommends the implementation of all predominant content forms. On this last point we could also detect a certain ambiguity. The fact of leaving each cataloger to decide whether a form of content is predominant or not, could result in the omission of items relevant to some extent. However, the

components that still continue to have some degree of ambiguity are the content qualifications of sensory specification, especially in their application in “Image” and “Object” content forms. The preliminary edition of Area 0 (“ISBD : International Standard Bibliographic Description: Area 0, Content Form and Media Type Area”), included an example that had both content forms: “Object + Image (olfactory) *Editorial comment:* a model and graphic (a 3-dimensional model of a coffee bean and 2-dimensional card with a micro fragrance-coated image of a coffee bean that can be scratched to produce the smell of coffee)”. After the publication of ISBD Consolidated Edition 2011, this example becomes obsolete due to two reasons. First, the content qualifications are now required whenever there are several possibilities for a given content form. Therefore, the term “visual” should be included in both forms of content. Secondly, the example does not include the media type, mandatory since the publication of ISBD Consolidated Edition 2011. An updated version of this example would then be: “Object (visual) + Image (still ; 2-dimensional; visual; tactile; olfactive) : unmediated” However, the publication of the full examples of the ISBD Consolidated (ISBD Examples Study Group) and ISBD Review Group) no longer includes this example, and only contains the most common types of documents (texts, electronic resources, music, cartographic materials, sound recordings). Does not include more complex resources as examples, such is the case of images and objects with sensory characteristics. Therefore, there are still uncertainties on the implementation of content qualifications of sensory specification in the description of these resources. Moreover, in the mapping of RDA Content Types to ISBD Content Forms and Media Types (Hostage) (provisional, subject to revision and last updated in May 2011) the application of content qualifications of sensory specifications is rather heterogeneous, which does not help to clarify these doubts either:

RDA Content Type	ISBD Content form
cartographic image	image (cartographic ; still; 2-dimensional ; visual)
cartographic moving image	image (cartographic ; movement ; 2-dimensional)
cartographic tactile image	image (cartographic ; still; 2-dimensional ; tactile)
cartographic tactile three-dimensional form	object (cartographic ; tactile)
cartographic three-dimensional form	object (cartographic ; visual)
still image	image (still ; 2-dimensional ; visual)
tactile image	image (still ; 2-dimensional ; tactile)
tactile three-dimensional form	object (tactile)
three-dimensional form	Object
three-dimensional form	image (still ; 3-dimensional)
three-dimensional moving image	image (movement ; 3-dimensional)
two-dimensional moving image	image (movement ; 2-dimensional)

Table 3

It would be important to define more specifically the application of content qualifications of sensory specification, according to the different content forms and also to improve and extend the full ISBD examples, including other types of more complex resources.

Mapping of RDA Content Types to ISBD Content Forms and Media Types (Hostage)

Table 4 also entitled *Complete RDA-ISBD mapping of content designators* (Dunsire, "Analysis of Content and Carrier Designators in the ISBD Consolidated Edition with Respect to the RDA-ONIX Framework") shows a mapping of all RDA content types (RDA section 6.9.1.3) to ISBD content forms and media types in Area 0. It was prepared by John Hostage, member of the ISBD Review Group and available on IFLA's website. This mapping is provisional, subject to revision and was last updated in May 2011. ISBD media types are included only when they are unambiguous.

An important difference between the ISBD Area 0 and vocabularies RDA is that Area 0 is composed by a syntax with specific punctuation to house each of its elements, in order to allow extensibility. The *Joint Steering Committee for Development of RDA* and the ISBD Review Group, met on 3-4 November 2011 to discuss the alignment between RDA and ISBD, and in particular between the vocabularies of the ISBD Area 0 and RDA vocabularies for content type and media type.

At this meeting, says Escolano Rodriguez ("Estado de la Cuestión de ISBD"), "it was stated that it is not possible to make all possible correspondences between ISBD as RDA vocabulary as proposed". Although both RDA and ISBD were based on ROF for the development of their vocabularies, "both standards have had differences in their results", and while the mapping performed by Hostage

RDA Content Type	ISBD Content form
cartographic image	image (cartographic ; still ; 2-dimensional ; visual)
cartographic moving image	image (cartographic ; moving ; 2-dimensional)
cartographic tactile image	image (cartographic ; still ; 2-dimensional ; tactile)
cartographic tactile three-dimensional form	object (cartographic ; tactile)
cartographic three-dimensional form	object (cartographic ; visual)
computer dataset	dataset : electronic
computer program	program : electronic
notated movement	movement (notated)
notated music	music (notated ; visual)
other (obsolete)	other content form
performed music	music (performed) : audio
Sounds	sounds : audio
spoken word	spoken word : audio
still image	image (still ; 2-dimensional ; visual)
tactile image	image (still ; 2-dimensional ; tactile)
tactile notated movement	movement (notated ; tactile)
tactile notated music	music (notated ; tactile)
tactile text	text (tactile)
tactile three-dimensional form	object (tactile)
Text	text (visual)
three-dimensional form	object
three-dimensional form	image (still ; 3-dimensional)
three-dimensional moving image	image (moving ; 3-dimensional)
two-dimensional moving image	image (moving ; 2-dimensional)

Table 4: Complete RDA-ISBD mapping of content designators

"shows some possible matches, it is not possible an absolute correspondence". "In this situation, during the meeting it was decided to make the RDF statement of this categorization and then to do mappings from each standard to the categorization RDA/ONIX". Therefore, the ISBD Review Group shall perform the correspondences of the elements of ISBD with categorization RDA / ONIX, to identify possible differences between them.

Moreover, on January 26th, 2012 the JSC, the Dublin Core Metadata Initiative (DCMI) Bibliographic Metadata Task Group (formerly DCMI/RDA Task Group), and ALA Publishing (on behalf of the co-publishers of RDA) announced the publication of a second set of vocabulary terms as linked open data in the Open Metadata Registry. These vocabularies include *Carrier type*,¹⁹ *Content type*²⁰ and *Media type*.²¹ At the moment, they are only published in English and in a preliminary version in German. As quoted in the announcement, Gordon Dunsire said: "These vocabularies are crucial for the selection and identification of information resources. Their publication as linked data in RDF allows the terms to be used by all bibliographic metadata communities in the Semantic Web environment. I look forward to the future development and publication of mappings from the vocabularies to the RDA/ONIX Framework. Similar mappings of other content and carrier vocabularies, such as those for ISBD area 0, will support metadata interoperability between communities and improve resource discovery for all" (Tillett, "Announcement: Publication of RDA Terms for Content, Carrier, Media Type Vocabularies").

¹⁹RDA Carrier Type: <http://metadataregistry.org/vocabulary/show/id/46.html>

²⁰RDA Content Type: <http://metadataregistry.org/vocabulary/show/id/45.html>

²¹RDA Media Type: <http://metadataregistry.org/vocabulary/show/id/37.html>

Relationship between ISBD and FRBR

The ISBD Consolidated Edition 2011 reflects the effort to bring description of all materials to the same state of conformity with the Functional Requirements for Bibliographic Records (FRBR). Therefore, for practical purposes it is necessary to accurately determine the alignment between ISBD and FRBR, both in relation to the attributes as to their status as mandatory or optional elements. In 2004, Tom Delsey conducted a mapping between the conceptual schema FRBR and ISBD standard entitled "Mapping ISBD Elements to FRBR Entity Attributes and Relationships." However, it was based on the specialized ISBDs, hence did not include, of course, the creation of Area 0. The ISBD Review Group raises, in the introduction to the ISBD Consolidated (ISBD Review Group and IFLA Cataloguing Section. Standing Committee, *ISBD : International Standard Bibliographic Description (Consolidated ed.)*), the need to update this document to specify the harmonic relationship between ISBD and FRBR. Moreover, "in trying to adapt ISBD terminology to the FRBR terms "work", "expression", "manifestation" and "item" and to replace terms such as "publication", the group encountered difficulties, owing in large part to the fact that the terms used in FRBR were defined in the context of an entity-relationship model conceived at a higher level of abstraction than the specifications for the ISBDs" (*ISBD : International Standard Bibliographic Description (Consolidated ed.)*). Taking into account the recommendation of the IME-ICC, the ISBD Review Group expressed not having performed a massive incorporation (or replacement) of the FRBR terminology in ISBD but rather tried to keep a specific terminology having in consideration certain elements to facilitate the future integration. For example, one change is the use of the term "resource" rather than "item" or "publication". This decision was taken in order to avoid confusion, because the use of the term "item" in the former ISBDs is different

from the term “item” as used in FRBR. The mapping conducted by Delsey (“Mapping ISBD Elements to FRBR Entity Attributes and Relationships”) showed “that many attributes that are present in ISBD but are not yet defined in FRBR, and also that each element in FRBR always has a correspondence to ISBD” (Bianchini and Guerrini, “From Bibliographic Models to Cataloging Rules: Remarks on FRBR, ICP, ISBD, and RDA and the Relationships Between Them”). In its introduction the FRBR document explains that “the basic elements of the model developed for the study (the entities, attributes, and relationships) were derived from a logical analysis of the data that are typically reflected in bibliographic records” and notes that “the principal sources used in the analysis included the ISBDs, the Guidelines for authority and reference entries (GARE), the Guidelines for Subject Authority and Reference Entries (GSARE), and the UNIMARC Manual” (International Federation of Library Associations and Institutions (IFLA) and IFLA Study Group on the Functional Requirements for Bibliographic Records). Therefore, “particular attention must be given to the fact that FRBR is a useful snapshot of the bibliographic universe obtained, mainly but not exclusively, of ISBD. This means that all objects (entities, relationships and attributes) extrapolated from FRBR have been identified within a description based on ISBD” (Bianchini and Guerrini, “Quis custodiet ipsos custodiet?: observaciones sobre la relación entre FRBR, ICP, ISBD y RDA”).

Delsey (“Mapping ISBD Elements to FRBR Entity Attributes and Relationships”) performed a mapping between elements of ISBD and FRBR but, as mentioned above, since in 2004 had not yet begun work on the ISBD Consolidated Edition, the ISBDs used for this analysis were specialized ISBDs. Therefore, the element analysed by Delsey that can be taken as an antecedent of Area 0 is the General Material Designation GMD:

ISBD Area	ISBD Element	FRBR Entity	FRBR Attribute/Relationship
0. Content form and media type	0.1 Content Form	3.2.1 Work	[Category of work]*There is no attribute currently defined in FRBR that corresponds to category of work
	0.2 Media Type	3.2.2 Expression	4.3.2 Form of expression: The form of expression is the means by which the work is realized (e.g., through alpha-numeric notation, musical notation, spoken word, musical sound, cartographic image, photographic image, sculpture, dance, mime etc.).
		3.2.3 Manifestation	[Category of carrier or medium] *There is no attribute currently defined in FRBR that corresponds to category of carrier or medium

Table 5

As stated by the ISBD Review Group, it is important to update this mapping to outline the compatibility between ISBD and FRBR, taking into account the FRBR entities and attributes, the mapping methodology and the new ISBD Consolidated 2011. An important element to consider in this analysis, based on the terminology of FRBR, is that “the ISBD is applied to describe manifestations, by means of description of the item in hand as an exemplar of the entire manifestation”. “In this way, the ISBD applies the Statement of International Cataloguing Principles, which establishes that “A bibliographic description typically should be based on the item as representative of the manifestation” (ISBD Review Group and IFLA Cataloguing Section. Standing Committee, *ISBD : International Standard Bibliographic Description (Consolidated ed.)*). In this sense, Escolano Rodriguez (“Estado de la Cuestión de ISBD”) notes that the

area 0 is "a first classification, contributes to a grouping by content and by media, thus serving as a bridge in the relationship established with FRBR entities Work/Expression, for content, and to merge with media type the full range of types of physical resources that meet the same characteristics (Manifestation)". However, the difficulty of this new mapping including Area 0, is the need to analyse each of the content forms, qualifications and media types in terms of FRBR and their specific correspondence with attributes of works, expressions and/or manifestations. An approach to this alignment between ISBD and FRBR can also be seen in the "LC CORE ELEMENTS FOR NOVEMBER GDR 2011" (Library of Congress, "LC RDA Core Elements for November 2011"), where an analysis was made of each element of RDA, its corresponding FRBR entity and its location in MARC. Although the scheme includes all elements of the RDA and does not specify their relationship to the ISBD, both could be related to the ISBD Area 0 indirectly, where the content type would be an attribute of expression and the media type would be an attribute of manifestation. However, further analysis is required to perform an affirmation in this regard.

RDA Element	RDA no.	FRBR entity	MARC encoding (tag only)
DESCRIBING CARRIERS			
<i>Media type</i>	3.2	Manifestation	337
<i>Carrier type</i>	3.3	Manifestation	338
IDENTIFYING WORKS AND EXPRESSIONS			
<i>Content type</i>	6.9	Expression	336

Table 6

Teaching of cataloging of non-book materials

For over 50 years cataloging has been found on bibliographic description according to an established typology of resources based on their carrier and/or nature. Both through specialized ISBDs and the different editions of AACR, the instructions for the description were organized on the basis of these broad categories. Due to the dynamic nature of library collections and the creation of new resources and the obsolescence of some formats, the categories were changed and reorganized over time.

Cataloging of non-book materials has historically been a world quite complex, compared with cataloging of printed books, with many special features to consider. While the structure of the record in ISBD is uniform to describe all materials, regardless of their nature or carrier, there are some elements of the description that are unique for these types of materials: the GMD in area 1 (until the creation of Area 0), Area 3 (used only for certain types of materials), SMD in Area 5 and especially the Area 7, for notes.

Quite evident proof of this organization of the bibliographic description is the structure of the first edition of the AACR, published in English (one American version and one English version) in 1967 and in Spanish in 1970. It was divided into three parts: I. Entry and Heading, II. Description and III. Non-book materials (Joint Steering Committee for Development of RDA). Later on, in the second edition of the AACR2 published in 1978, the assignment of access points was moved to the second part and the description of all materials was grouped in the first part. This last was, at the same time, divided into a general section and in 9 chapters for the individual formats. Subsequently, three reviews were published in 1988, 1998 and 2002 including new amendments. The latest version of AACR2,

used until present days in libraries around the world, includes the subdivision of the first part in 13 chapters based on their nature, carrier or media type: books, pamphlets and printed sheets, cartographic materials, manuscripts, music, sound recordings, films and video recordings, graphic materials, electronic resources, three-dimensional artifacts and realia, microforms, continuing resources and analysis.

Also, the ISBD has been originated in the organization of cataloging according to different natures of documents, carriers or media types, as evidenced in the existence of 8 ISBDs prior to the creation of the ISBD Consolidated: ISBD (G) -general-, ISBD (CM) -cartographic materials-, ISBD (S) -serials-, ISBD (NBM) -non bibliographic materials-, ISBD (A) -antiquarian-, ISBD (PM) -printed music-, ISBD (M) -monographs- and ISBD (CF) -computer files- then updated for ISBD (ER) -electronic resources-. Nowadays, international research in cataloging, carried out by the Joint Steering Committee for Development of RDA, the ISBD Review Group, the IFLA Study Group on FRBR, IFLA Cataloguing Section and the Library of Congress are changing this paradigm. Both the FRBR model and the new Consolidated ISBD as RDA structure, make it clear that it is no longer about the carrier, but the entities, their attributes and the relationships between them. However, we find ourselves in a time of a change so profound and so rapid that our programs of study still do not reflect, in many cases, this paradigm shift. Balance must be found in order to bring this new axis to the classrooms but still providing practical tools for professional cataloging. The theoretical and conceptual elements are extremely important, and involve a deep reflection and analysis to enable better description and more efficient access to library resources. In this respect, Bianchini and Guerrini ("Quis custodiet ipsos custodiet?: observaciones

sobre la relación entre FRBR, ICP, ISBD y RDA”) reflect and affirm: "While our logical model of the bibliographic universe has changed completely, has not been abandoned, nor the Paris Principles, nor the AACR (...) No one will give up the Principles of Paris or AACR before having thoroughly tested the new proposals. This should also apply to ISBD, especially because it remains the most recognized standard and more popular worldwide and it is in the concrete base of FRBR". We are in the midst of a rapidly changing conjuncture, a time of transition and new paradigms. This will generate concerns, fears and difficulties in teaching and working with different standards. Cataloging teachers have a huge challenge: to integrate all these new elements in teaching students trying to assimilate the foundations and principles of cataloging and simultaneously incorporate new schemes to be prepared for the future. But we must not forget that not only do we offer a broad view of the bibliographic universe, allowing them to look critically and participate in the construction of the evolution of these investigations, but also practical tools for use today as cataloging librarians.

Works cited

- Bianchini, Carlo and Mauro Guerrini. "From Bibliographic Models to Cataloging Rules: Remarks on FRBR, ICP, ISBD, and RDA and the Relationships Between Them". *Cataloging & Classification Quarterly* 47.2. (2009): 105–124. <http://eprints.unifi.it/archive/00002177/01/CCQ_Bianchini-Guerrini_From_907343194-1.pdf>. (Cit. on p. 39).
- . "Quis custodiet ipsos custodiet?: observaciones sobre la relación entre FRBR, ICP, ISBD y RDA". *Anales de Documentación* 12. (2009): 32–341. <http://eprints.unifi.it/archive/00001840/01/Quis_custodiet_spagnolo_Pdf.67871.pdf>. (Cit. on pp. 21, 39, 43).
- . "Quis custodiet ipsos custodiet?: Osservazioni sulle Relazioni fra FRBR, ICP, ISBD e RDA". *Il bibliotecario* 3.1. (2009): 59–85. <http://eprints.unifi.it/archive/00001743/01/Quis_custodiet_ipsos_custodes.pdf>.
- Chiverto, Daniel Salamanca. "La Categorización de Contenidos y Medios en la Descripción Bibliográfica: la Designación General de Material (DGM) y su Pasado, Presente y Futuro en la Era Digital". *Revista Española de Documentación Científica* 31.4. (2008): 527–551. <<http://redc.revistas.csic.es/index.php/redc/article/download/442/454>>. (Cit. on pp. 3–5, 8).
- Delsey, Tom. "Mapping ISBD Elements to FRBR Entity Attributes and Relationships". (2004). <<http://www.ifla.org/VII/s13/pubs/ISBD-FRBR-mappingFinal.pdf>>. (Cit. on p. 39).
- . "The Logical Structure of the Anglo-American Cataloguing Rules, Part I". (1998). <<http://www.rda-jsc.org/docs/aacr.pdf>>.
- Dunsire, Gordon. "Analysis of Content and Carrier Designators in the ISBD Consolidated Edition with Respect to the RDA-ONIX Framework". Proc. of Seventy-Sixth World Library and Information Congress, Gothenburg, 2010. (Cit. on pp. 29–31, 35).
- . "Distinguishing Content from Carrier: The RDA/ONIX Framework for Resource Categorization". *D-Lib Magazine* 13.1-2. (2007). <<http://www.dlib.org/dlib/january07/dunsire/01dunsire.html>>. (Cit. on p. 29).
- . "Mapping ISBD and RDA Element Sets; Briefing/discussion Paper (6JSC/Chair/4)". (2011). <<http://www.rda-jsc.org/docs/6JSC-Chair-4.pdf>>.
- . "Mapping ISBD Area 0 Vocabularies to RDA Carrier and Content Vocabularies Via the RDA/ONIX Framework for Resource Categorization (ROF): Briefing/discussion Paper (6JSC/Chair/5)". (2011). <<http://www.rda-jsc.org/docs/6JSC-Chair-5.pdf>>.
- . "UNIMARC, RDA and the Semantic Web (Meeting No. 105)". Proc. of Seventy-fifth IFLA General Conference, Milan, 2009.

- Federación Internacional de Asociaciones de Bibliotecas e Instituciones (IFLA) and Comisión de Traducción de la Biblioteca Nacional (España), eds. "Descripción Bibliográfica Internacional Normalizada (ISBD) (Ed. preliminar consolidada)". (2008). <<http://www.bne.es/es/Servicios/NormasEstandares/ISBDconsolidada/Docs/ISBDconsolidada.pdf>>.
- , eds. "Descripción Bibliográfica Internacional Normalizada (ISBD): Área 0, Área de Forma del Contenido y de Tipo de Medio". (2009). <http://www.bne.es/opencms/es/Servicios/NormasEstandares/ISBDconsolidada/Docs/ISBD_Area_0.pdf>.
- Federación Internacional de Asociaciones de Bibliotecas e Instituciones (IFLA) and Elena Escolano Rodríguez. "Declaración de Principios Internacionales de Catalogación". (2010). <<http://www.bne.es/es/Servicios/NormasEstandares/Docs/principioscatalogacion2009.pdf>>.
- Grupo de Estudio de la IFLA sobre los Requisitos Funcionales de los Registros Bibliográficos, Xavier Agenjo, and María Luisa Martínez-Conde. "Requisitos Funcionales de los Registros Bibliográficos (FRBR)". (2004). <Availableonline: <http://www.ifla.org/VII/s13/frbr/frbr-es.pdf>>.
- Guerrini, Mauro. "GMD: its Function and its History". *Cataloging & Classification Quarterly* 38.2. (2004): 61–74. (Cit. on pp. 4, 5).
- Hostage, John. "Mapping of RDA Content Types to ISBD Content Forms and Media Types". (2011). <<http://www.ifla.org/node/5618>>. (Cit. on pp. 33, 35).
- Huthwaite, Ann. "Class of Materials Concept and GMDs (4JSC/ACOC rep/1)". (2003). <http://www.d-nb.de/standardisierung/pdf/papers_huthwaite.pdf>. (Cit. on p. 6).
- International Federation of Library Associations and Institutions (IFLA). "Superseded ISBD". (2009). <<http://www.ifla.org/node/900>>.
- International Federation of Library Associations and Institutions (IFLA) and IFLA Study Group on the Functional Requirements for Bibliographic Records. "Functional Requirements for Bibliographic Records: Final Report / Approved by the Standing Committee of the IFLA Section on cataloguing. September 1997; as amended and corrected through February 2009". (2009). <http://www.ifla.org/files/cataloguing/frbr/frbr_2008.pdf>. (Cit. on p. 39).
- ISBD Examples Study Group) and ISBD Review Group. "Full ISBD Examples: Supplement to the Consolidated Edition of the ISBD: International Standard Bibliographic Description". (2011). <http://www.ifla.org/files/cataloguing/isbd/isbd-examples_2011.pdf>. (Cit. on p. 33).
- ISBD Review Group. "Report of Activities, 2010-2011". (2011). <http://www.ifla.org/files/cataloguing/isbdrg/activities_2010-2011.pdf>.
- , "Report of Activities, August 2010-June 2011". (2011). <http://www.ifla.org/files/cataloguing/isbdrg/isbd-xml-activities_2010-2011.pdf>.

- ISBD Review Group and IFLA Cataloguing Section. Standing Committee. "ISBD : International Standard Bibliographic Description: Area 0, Content Form and Media Type Area". (2009). <http://www.ifla.org/files/cataloguing/isbd/area-0_2009.pdf>. (Cit. on pp. 1, 2, 7, 8, 33).
- . *ISBD : International Standard Bibliographic Description (Consolidated ed.)* IFLA Series on Bibliographic Control 31. 2010. (Cit. on pp. 2, 10, 38, 40).
- . *ISBD : International Standard Bibliographic Description (Consolidated ed.)* IFLA Series on Bibliographic Control 44. Berlin, München: De Gruyter Saur, 2011. (Cit. on pp. 10, 32, 38).
- . "ISBD : International Standard Bibliographic Description (Preliminary Consolidated ed.)" (2009). <http://archive.ifla.org/VII/s13/pubs/ISBD_consolidated_2007.pdf>. (Cit. on pp. 3, 4).
- Joint Steering Committee for Development of RDA. "A Brief History of AACR: Anglo-American Cataloguing Rules". (2009). <<http://www.rda-jsc.org/history.html>>. (Cit. on p. 42).
- Joint Steering Committee for Development of RDA and Joint Steering Committee for Revision of AACR. "RDA/ONIX Framework for Resource Categorization (Version 1.0, released August 1) (5JSC/Chair/10)". (2006). <<http://www.rda-jsc.org/docs/5chair10.pdf>>.
- Joint Steering Committee for Revision of AACR. "Designación General del Material". *Reglas de Catalogación Angloamericanas (2ª ed., rev. 2002, act. 2003)*. Trans. Margarita Amaya de Heredia. Bogotá: Rojas Eberhard, 2004. 1–12.
- Library of Congress. "LC RDA Core Elements for November 2011". (2011). <http://www.loc.gov/aba/rda/pdf/core_elements.pdf>. (Cit. on p. 41).
- . "Network Development and MARC Standards Office (2007). MARC 21 Format for Bibliographic Data". (2011). <<http://www.loc.gov/marc/bibliographic/ecbdhome.html>>.
- Library of Congress and Oficina de Desarrollo de Redes y Normas MARC. "Formato MARC21 Conciso para Datos Bibliográficos". (2007). <<http://www.loc.gov/marc/bibliographic/ecbdspa.html>>.
- Pape, Philippe Le. "FRBR en UNIMARC: Propositions du Cfu Validées par le PUC". Proc. of PUC Meeting, Lisbon. 2011. (Cit. on p. 23).
- Permanent UNIMARC Committee. "Minutes of the 21st Meeting". (2010). <http://www.ifla.org/files/uca/puc-meetings/21_PUC_meeting_%20report_2010.pdf>.
- "RDA/ONIX Framework for Resource Categorization". (2006). <<http://www.collectionscanada.ca/jsc/docs/5chair10.pdf>>. (Cit. on p. 29).
- Rodríguez, Elena Escolano. "Edición Preliminar de la ISBD Consolidada". Proc. of III Encuentro Internacional de Catalogadores, Buenos Aires. 2007. (Cit. on p. 8).

- . “Estado de la Cuestión de ISBD”. Proc. of VII Encuentro Internacional y III Nacional de Catalogadores, Buenos Aires. 2011. (Cit. on pp. 9, 35, 40).
- Rodríguez, Elena Escolano and Ageo García, eds. “Declaración de Principios adoptados por la Conferencia Internacional sobre Principios de Catalogación, París, Octubre de 1961”. (2005). <http://www.bne.es/es/Servicios/NormasEstandares/Docs/Paris_1961.pdf>.
- Rodríguez, Elena Escolano and Dorothy McGarry. “International Standard Bibliographic Description: Updating the Consolidated ISBD”. *International Cataloging & Bibliographic Control* 38.1. (2009): 13–14.
- Rodríguez, Elena Escolano, et al. “Novedades de ISBD (Meeting No. 107)”. Proc. of Seventy-fifth IFLA General Conference, Milan. 2009.
- Tillett, Barbara. “Announcement: Publication of RDA Terms for Content, Carrier, Media Type Vocabularies”. (2012). <<http://www.mail-archive.com/rda-listserv.lac-bac.gc.ca/msg06558.html>>. (Cit. on p. 37).
- . “General Material Designations (GMDs) (4JSC/Chair/73)”. (2001). <<http://www.rda-jsc.org/docs/gmd.pdf>>. (Cit. on pp. 3, 4, 6).
- Weihs, Jean. “General Material Designation in the Twenty-First Century: Results of a Survey”. (2001). <http://www.olacinc.org/drupal/capc_files/gmd.pdf>.

MARIA VIOLETA BERTOLINI, Biblioteca del CEI y Centro de Referencia de la OMC, Buenos Aires.

violebertolini@gmail.com

Website: <http://ar.linkedin.com/in/mariavioletabertolini>

Bertolini, M. V. "The new ISBD Area 0 and the teaching of cataloging of non-book materials". *JLIS.it* Vol. 3, n. 1 (Giugno/June 2012): 4794-1-4794-49. DOI: [10.4403/jlis.it-4794](https://doi.org/10.4403/jlis.it-4794). Web.

ABSTRACT: The International Standard Bibliographic Description (ISBD) Area 0 was first included in the ISBD Consolidated Edition published in July 2011, as a replacement of the General Material Designation (GMD). This work briefly makes reference to the antecedents of area 0 and to the reasons that lead to its creation. Also, it details its structure, elements and examples of use and analyzes its compatibility with the MARC format, the RDA vocabularies and with FRBR model. Finally, in relation to the creation of this new area and to its application in the diversity of existing bibliographical resources, it reflects on the change of paradigm of cataloguing and the challenges in the teaching of cataloguing of non-book materials.

KEYWORDS: Cataloguing; FRBR; ISBD; MARC; Non-book materials; RDA

ACKNOWLEDGMENT: This article has been presented at the Conference "VII Encuentro Internacional y III Nacional de Catalogadores", held in the Biblioteca Nacional de la República Argentina, in Buenos Aires (AR) on 23-25 November 2011. The Conference paper is available online at <http://www.bn.gov.ar/vii-encuentro-de-catalogadores>.

Submitted: 2012-02-12

Accepted: 2012-03-02

Published: 2012-03-02





Il progetto National Bibliography Number Italia (NBN:IT). Un identificatore persistente a supporto del deposito legale nazionale delle risorse digitali

Emauele Bellini, Chiara Cirinnà, Maurizio Lunghi,
Cinzia Luddi, Maurizio Messina, Giovanni Bergamin,
Raffaele Messuti, Giovanna Cordani, Roberto
Delle Donne, Francesca Rossi, Marialaura Vignocchi,
Stefania Arabito, Jordan Piščanc'

Introduzione

L'associazione di un Persistent Identifiers (PI) a una risorsa digitale è considerata oggi una best practice che contribuisce alla risoluzione del problema del reperimento affidabile delle risorse in rete, andando ben oltre la soluzione dell'errore Internet 404. Attualmente i PI sono considerati un mezzo per poter garantire non solo l'identificazione e l'accesso nel lungo periodo, ma anche la certificazione, l'autorevolezza e l'affidabilità delle risorse stesse a patto però che l'implementazione dell'infrastruttura prenda in considerazione non



solo i requisiti tecnici, ma soprattutto quelli organizzativi, di processo, di comunità, di conservazione etc. L'associazione di un PI a una risorsa digitale può essere usata come un concreto strumento di localizzazione affidabile, ma, allo stesso tempo, può supportare la certificazione dell'autenticità della risorsa, della sua provenienza, dei suoi diritti di sfruttamento etc.

Ovviamente la garanzia di una reale persistenza degli identificativi è principalmente legata all'impegno mostrato da un'organizzazione nell'assegnarli e gestirli. Questo impegno deve declinarsi nell'adozione di precise politiche di servizio, nell'implementazione di specifiche soluzioni tecnologiche, nella definizione di un modello di business sostenibile, nell'identificazione della comunità di riferimento, nel rispetto di opportuni standard condivisi internazionalmente. La reale persistenza di un sistema di identificazione si deve quindi basare sul rispetto di una serie di criteri qualitativi e quantitativi verificabili così come proposto dai risultati di iniziative internazionali, quali NESTOR¹ («NESTOR Catalogue of criteria for assessing the trustworthiness of PI systems») e APARSEN² (Bazzanella et al.).

Attualmente sono disponibili diverse tecnologie e standard per l'implementazione di un sistema di attribuzione di PI; manca tuttavia un accordo sulla loro adozione, anche perché alcuni di questi sistemi sono nati come soluzioni tecniche senza il supporto delle comunità di utenti che si aspettano dai PI livelli di servizio. Siste-

¹NESTOR, <http://www.langzeitarchivierung.de/eng>, è un network di competenza tedesco sulla digital preservation e ne fanno parte biblioteche, archivi, musei e esperti del settore.

²Alliance for Permanent Access to the Records of Science Network, <http://www.alliancepermanentaccess.org/index.php/current-projects/aparsen>, è un network of excellence sulla digital preservation finanziata dalla commissione europea che si occupa di studiare tutti gli aspetti della conservazione dei record scientifici, dagli standard alla loro identificazione, dalle soluzioni di storage alla creazione di un centro di competenza virtuale europeo.

mi come il PURL³ oppure i Cool URI⁴ (Berners-Lee) presentano indubbiamente notevoli vantaggi (il più rilevante tra questi è la loro immediata dereferenziabilità⁵ attraverso il protocollo HTTP⁶), ma il limite di queste soluzioni è che la loro persistenza non è in principio garantita da una terza parte indipendente e autorevole. Inoltre, nell'ambito del progetto europeo APARSEN è stata condotta un'indagine i cui risultati mostrano che un numero limitato di approcci, come il Digital Object Identifier DOI (33%), il sistema Handle (29%) e gli identificatori URN-based (Moats) (25%), sembrano ormai prevalere a livello europeo (Bazzanella et al.).

Una particolare attenzione merita in ogni caso il Cool URI. Com'è noto il Cool URI è un approccio alla persistenza delle URL basato sulla loro progettazione. Questo approccio principalmente tecnico è considerato una best practice dal W3C per l'implementazione delle URI del semantic web e dei Linked Data.⁷ L'assunzione di base è che una progettazione corretta delle URI dovrebbe ridurre i casi in cui risulterà necessario cambiarle, garantendone la loro stabilità nel tempo. Un esempio di questa best practice è quello di evitare di esplicitare l'estensione delle pagine web come .php o .asp, in modo che se cambia la tecnologia di base (da PHP ad ASP) la URI rimane

³OCLC Persistent URL è un servizio che consente agli utenti di generare delle URL con certe caratteristiche di persistenza riferendo a queste le URL delle proprie risorse considerate più soggette a cambiamenti, migrazioni, ecc. URL <http://purl.oclc.org>.

⁴Cool URI è una best-practice del W3C che intende rendere persistenti le URL tramite l'adozione di una serie di regole e pratiche di gestione in grado di ridurre i casi in cui risulterebbe necessario cambiarle o dismetterle, <http://www.w3.org/Provider/Style/URI.html>.

⁵La locuzione "dereferenziare una URI" è riferita alla capacità di recuperare la rappresentazione della risorsa identificata da quella URI tramite una applicazione (es. browser). In altre parole è la capacità di accedere ad esempio a una pagina web, o a un PDF all'intero di un browser tramite una URL.

⁶<http://www.ietf.org/rfc/rfc2616.txt>.

⁷<http://www.w3.org/TR/2007/WD-cooluris-20071217>, <http://www.w3.org/DesignIssues/LinkedData.html>.

la stessa. In questa ottica, la persistenza si baserebbe esclusivamente sull'impegno della singola istituzione; s'instaura così una relazione di trust diretta tra l'utente e l'istituzione stessa, senza mediazioni. In questo senso il trust è il livello di fiducia che un utente ha nella capacità di un'istituzione di mantenere le caratteristiche dichiarate del servizio offerto. Il livello di fiducia viene ciclicamente verificato durante l'interazione che l'utente ha con il servizio stesso, e dove si ha una conferma positiva, il trust si rafforza; in caso contrario decresce. È ben noto, purtroppo, che l'impegno della singola istituzione non è più sufficiente a garantire né la persistenza della URL né tanto meno la certificazione delle risorse in termini di provenienza, autenticità, integrità, conservazione etc. Nella pratica, le risorse si muovono sulla rete, vengono modificate o cancellate a causa di una infinità di fattori che non possono essere sempre predeterminati dalle politiche di gestione dei contenuti delle singole istituzioni, né tanto meno governate da best practice esclusivamente tecniche. Un caso tipico è quando un'istituzione chiude le sue attività perché viene comprata da un'altra istituzione o viene soppressa, oppure semplicemente cambia nome. In questi casi gli oggetti possono essere rinominati per essere adattati al workflow interno della nuova istituzione, o ceduti ad altre istituzioni, o nel peggiore dei casi cancellati perché non più corrispondenti agli obiettivi istituzionali. È evidente che tutte queste azioni possono produrre l'invalidazione delle vecchie URL indipendentemente da come queste siano state costruite. Questo potrebbe non essere un problema se l'istituzione non gestisce contenuti di valore scientifico, culturale o amministrativo, ma cosa diversa è se la medesima sorte la subiscono istituzioni come editori scientifici, biblioteche, archivi, etc. In questi casi, ad esempio, bibliografie basate su semplici URL o anche Cool URI, facenti riferimento a risorse che erano presenti negli archivi di queste istituzioni, non potranno più essere usate, ad esempio, per verificare il lavoro scien-

tifico o per calcolare indici bibliometrici. La cosa più critica, però, è l'impossibilità di implementare sistemi di verifica dell'autenticità, della provenienza e dell'integrità di queste risorse in assenza di una terza parte che si faccia garante dell'associazione nome – risorsa.

L'uso dei PI consente, quindi, di inserire non solo nel web ma anche nel nuovo web semantico quel necessario livello di trust che attualmente manca allo scenario così come auspicato dai risultati del seminario sui PI e Linked Open Data confluiti nel Den Haag Manifesto,⁸ a cui la Fondazione Rinascimento Digitale ha contribuito attivamente. Questo livello di trust è implementato attraverso l'identificazione di una terza parte indipendente e autorevole (come ad esempio le Biblioteche Nazionali) con un preciso mandato che tenga in considerazione non solo i requisiti tecnici, ma anche quelli di comunità (ad esempio la granularità dell'identificativo), organizzativi (autorevolezza della Registration Authority riconosciuta nel dominio di applicazione, impegno sulla durata del servizio) e di processo (policy di dominio sulla selezione delle risorse, etc.) (Bellini et al., «Semantics-Aware Resolution of Multi-part Persistent Identifiers»; «Persistent Identifiers distributed system for Cultural Heritage digital objects»).

Il progetto NBN Italia

Il progetto NBN Italia (NBN:IT) ha inteso realizzare un servizio di identificazione persistente basato sullo standard NBN che assegna alle biblioteche nazionali la responsabilità esclusiva della gestione

⁸http://www.knowledge-exchange.info/Admin/Public/DWSDownload.aspx?File=%2fFiles%2fFiler%2fdownloads%2fPersID%2fWorkshop+POID%2fDen+Haag+Manifesto+20110825_2.pdf. Si noti in ogni caso la differenza tra l'identificatore di tipo URL appena citato e gli identificatori riportati alle precedenti, costruiti come Cool URI.

del sottodominio nazionale NBN:IT. Partendo dall'attuale orientamento per i sistemi NBN, derivato da progetti come PersID, da report come CENL Task Force on Persistent Identifiers e dalle raccomandazioni contenute ne The National Libraries Resolver Discovery Service («CENL Task Force on Persistent Identifiers, Report 2007»), dove se ne sottolinea il ruolo a supporto della digital preservation, in Italia si è consolidata l'idea di legare NBN:IT a un servizio che è istituzionalmente affidato alle biblioteche nazionali: il deposito legale. I primi studi intorno al NBN sono cominciati nel 2007 su iniziativa della Fondazione Rinascimento Digitale con il prezioso supporto dell'Università Statale di Milano con significativi risultati dal punto di vista della ricerca (Bellini et al., «Persistent Identifiers distributed system for Cultural Heritage digital objects»; «Semantics-Aware Resolution of Multi-part Persistent Identifiers»; Bellini e Ceravolo). Le attività sono poi proseguite con una prima fase di sperimentazione che ha visto lo sviluppo dell'architettura e il test di vari prototipi (Bellini et al., «Persistent Identifier Distributed System for Digital Libraries, Information Technology»; «Persistent identifiers: jNBN, a JEE application for the management of a national NBN infrastructure»). Infine, in considerazione della vocazione di NBN verso la digital preservation, il progetto NBN:IT è confluito nell'ambito del Deposito Legale nazionale, diventandone oggi parte integrante. L'iniziativa NBN vede oggi coinvolti il consorzio di Magazzini Digitali (Vignocchi, Bergamin e Messuti; «Le linee guida per l'accesso aperto alle tesi di dottorato») per il deposito legale composto dalla Biblioteca Nazionale Centrale di Firenze (BNCF), la Biblioteca Nazionale Centrale di Roma (BNCR), la Biblioteca Nazionale Marciana di Venezia e la Fondazione Rinascimento Digitale (FRD) e la CRUI che sovrintende la sperimentazione con le tesi di dottorato. Attualmente il progetto NBN è in fase avanzata di sviluppo e i suoi principali obiettivi possono essere riassunti nei seguenti punti:

1. la creazione di un registro nazionale di nomi stabile e affidabile come strumento a supporto del deposito legale per la certificazione degli oggetti digitali prodotti dalle istituzioni culturali, scientifiche e educative,
2. l'accesso permanente alle risorse digitali prodotte dalle istituzioni culturali italiane, incluso il materiale digitalizzato o non ancora pubblicato,
3. la condivisione di politiche per la conservazione di lungo periodo degli oggetti digitali,
4. la sostenibilità dei costi del servizio e le responsabilità di gestione dei nomi da parte delle istituzioni.

NBN: un identificatore a supporto del deposito legale

La decisione di associare un servizio di identificazione persistente a un servizio di deposito legale ha delle implicazioni importanti, che meritano qualche considerazione. La definizione del nuovo servizio NBN prende avvio dalle considerazioni espresse nel report CENL Task Force on Persistent Identifiers [2] e nelle CENL Recommendation - The National Libraries Resolver Discovery Service (RDS) («CENL Task Force on Persistent Identifiers, Report 2007») e si consolida nell'idea di collegare NBN a un servizio specifico, andando ben oltre la sola esigenza di identificazione e realizzando un sistema a valore aggiunto. Già nel 2007 i report del CENL promuovevano l'adozione di NBN da parte delle Biblioteche Nazionali, scoraggiando l'uso di altri sistemi. Inoltre, l'idea che NBN fungesse anche da risolutore di ultima istanza di DOI, una volta che una risorsa digitale non fosse stata più accessibile dall'archivio dell'editore, ci fa comprendere la differenza e la complementarità di NBN rispetto agli altri sistemi. L'iniziativa italiana ha recepito in pieno queste

indicazioni spingendo le implicazioni di queste scelte ben oltre la sfera applicativa. Infatti, il conferimento all'identificativo NBN di significati ulteriori, a nostro avviso, consente di comunicare quel valore di "fiducia" atteso dagli utenti dei servizi erogati dalle pubbliche amministrazioni in generale, e, in particolare, dei servizi di conservazione e identificazione.

Il ruolo iconico di NBN

Un identificatore persistente utilizzato da organismi istituzionali per identificare un'entità o un oggetto prodotto dall'organizzazione stessa come risultato di un processo amministrativo e burocratico ha un potenziale ruolo iconico nella definizione dei requisiti di trust nell'information society («Final Report of the Project Team of CEN/ISSS eGovernment Focus Group, CEN»; Winer; Wilson). Luhmann («Trust: a mechanism for the reduction of social complexity») definisce il trust come un meccanismo per ridurre l'incertezza, la complessità e il rischio che percepiscono gli utenti riguardo all'entità in esame. In quest'ottica, quindi, gli utenti si aspettano che l'output di un sistema ritenuto affidabile, sia corretto e in linea con le loro aspettative. La reiterazione di questa conferma contribuisce a stabilizzare questa fiducia. L'attribuzione di un identificativo a una risorsa, assegnato da parte di una pubblica amministrazione, dovrebbe implicare il superamento da parte della risorsa stessa di un processo ufficiale di riconoscimento (carta d'identità), di presa in carico (numero di protocollo), di tracciamento (codice fiscale, targa), e così via. La possibilità di verifica dell'avvenuto processo è garantita non solo dal mantenimento in appositi registri (anagrafe, protocollo, etc.) dell'associazione tra l'identificativo assegnato e l'entità oggetto del processo, ma anche dalla loro immediata accessibilità a fini di controllo. Quindi, la sola presenza di un identificativo di questo tipo conferisce all'entità a cui è assegnato, un livello di trust supe-

riore rispetto a entità della stessa natura che ne sono sprovviste. Il ruolo iconico di un identificatore deriva, quindi, dal fatto che la sua assegnazione (o adozione) è subordinata al superamento di un processo le cui modalità di attuazione sono conosciute, accettate e formalmente e/o legalmente disciplinate. In accordo a questa visione si è ritenuto inevitabile dotare il processo di deposito legale di un sistema di identificazione delle risorse stesse. L'associazione di un NBN a una risorsa certificherà pertanto all'utente che quella stessa risorsa è stata legalmente depositata (o che è in corso il deposito legale) in accordo alle regole che disciplinano questo processo. La dimostrazione dell'avvenuto processo può essere perpetrata solo tramite la risoluzione e confermata dall'utente finale.

Il rapporto tra NBN e DOI/Handle

Dopo un'iniziale proliferazione di tecnologie e proposte di standardizzazione, attualmente il dominio sembra pragmaticamente focalizzato su soluzioni come DOI, Handle e identificativi URN based come NBN. I sistemi tra loro non sono comunque equivalenti. Infatti, sia Handle che DOI (che si basa su l'infrastruttura Handle) possono essere assegnati a qualunque tipo di risorsa e la loro funzione principale è quella di identificazione (per Handle) o di indirizzamento alle pagine degli editori dove è possibile acquistare la risorsa (per DOI). Nessuno dei due però può dirsi strettamente subordinato a un workflow/processo specifico e certificato come avviene ad esempio per l'assegnazione di ISBN (processo di pubblicazione) o per il rilascio della carta di identità (processo di riconoscimento). Infatti, anche se il DOI è generalmente usato dagli editori nella fase finale del processo di pubblicazione, chiunque può acquisire tale identificativo commerciale per le proprie risorse senza che queste ultime siano sottoposte necessariamente a un processo di selezione, catalogazione, archiviazione, etc. Il DOI, come Handle, è quindi un

identificativo “generico” (avere un DOI o un Handle di per sé, non dice niente di più del fatto che la risorsa sia identificata su Internet) e pertanto supporta l’erogazione di altri tipi di servizi come quelli citazionali, che non sono in alternativa a quelli forniti attraverso NBN, ma vi si aggiungono, completando l’offerta. Infine, sia per il DOI che per Handle, la responsabilità della stabilità delle risorse e della loro conservazione è interamente affidata al content provider, quindi, nel caso l’istituzione non abbia adottato specifiche politiche di conservazione, il rischio che nel lungo periodo il DOI non risolva più correttamente è concreto. Diversamente, NBN associato al servizio di deposito legale assicura un’affidabilità maggiore in quanto beneficia dell’assunzione di responsabilità della pubblica amministrazione di conservare nel lungo periodo le risorse. In quest’ottica, la garanzia di risoluzione di un NBN nel lungo termine è data dal fatto che questo punterà sempre almeno alla copia conservata nel deposito legale.

Attualmente le risorse possono già ricevere più di un identificativo durante il loro ciclo di vita. Ad esempio, un articolo pubblicato su una rivista non Open Access e depositato in un repository istituzionale OA basato su un sistema DSpace il cui dominio è registrato al CNRI, avrà come minimo un ID interno OAI (non risolvibile), un Handle (risolvibile alla risorsa depositata nel repository) e un DOI (risolvibile principalmente alla copia depositata presso l’editore), quindi l’introduzione di un nuovo identificativo con le caratteristiche precedentemente descritte non potrà che rappresentare un valore aggiunto per la risorsa stessa. Infatti, questi identificativi, tutti gestiti dai metadati descrittivi associati alla risorsa, di per sé non sono in contrasto in quanto, come abbiamo detto, assolvono a servizi diversi, e il NBN rappresenta un metadato aggiuntivo in grado di informare l’utente che esiste una copia della risorsa legalmente depositata alla biblioteca nazionale. La prova del deposito è

ottenibile tramite il servizio di risoluzione del nome.

Lo standard IETF RFC 3188 – National Bibliography Number

Il PI definito come National Bibliography Number (NBN) (Hakala) nasce nel contesto del ruolo e dalle responsabilità che le biblioteche nazionali hanno nel campo del controllo bibliografico universale (in particolare con le bibliografie nazionali).⁹ Si tratta di uno standard basato su Uniform Resource Name (URN) (Masinter e Sollins; Moats; Daigle et al.) e definito dall’RFC 3188 in ambito IETF. Lo standard NBN riprende molti degli aspetti definiti nell’RFC 2288 (Lynch, Preston e Daniel) nel quale si fa riferimento all’utilizzo degli identificatori bibliografici come l’International Standard Book Number (ISBN), l’International Standard Serials Number (ISSN) («NISO/ANSI/ISO 2108:1992 Information and documentation - International standard book number (ISBN)»; «ISO 3297:1986 Documentation - International standard serial numbering (ISSN)»), e il Serial Item and Contribution Identifier (SICI) («NISO/ANSI Z39.56-1997 Serial Item and Contribution Identifier»). I namespace basati sullo standard URN (quindi anche NBN) sono fortemente orientati alla sintassi dei nomi e alle responsabilità della gestione dei sub-namespace. Infatti, esistono diversi gradi di libertà per l’implementazione del servizio, che può essere diverso per comunità diverse. Nel caso di NBN, lo standard prevede una ripartizione di responsabilità simile ai domini DNS dove è previsto un suffisso che identifica il paese di riferimento (es. NBN:IT:xxxx oppure NBN:DE:yyyy).

Quindi la struttura tipica di un NBN è la seguente:

URN:NBN:<country code>:<sub-namespace>-<assigned NBN>

⁹Per il concetto di Controllo Bibliografico Universale si veda Solimine (*Controllo bibliografico universale*).

dove "country code" è il codice del Paese, indicato secondo lo standard ISO 3166, il "sub-namespace" è il codice del sottodominio e "assigned NBN" è la stringa di caratteri scelta.

La responsabilità per la gestione del sottodominio nazionale è delegata, come previsto dallo standard, alle biblioteche nazionali di ogni paese (Hakala). NBN è stato registrato al registro Internet Assigned Numbers Authority (IANA) degli URN nell'ambito del Nordic Metadata Project su proposta della Conference of Directors of National Libraries (CDNL) e Conference of European National Libraries (CENL). Il primo obiettivo dello standard era quello di avere un sistema di identificazione per tutte quelle risorse digitali che non possedevano un identificativo perché fuori da circuiti di produzione specifici, come avviene invece per i libri con ISBN o i periodici con ISSN. Esempi di queste risorse che andavano a collocarsi nella cosiddetta letteratura grigia sono le tesi di dottorato, i rapporti tecnici, gli oggetti digitalizzati come i manoscritti antichi etc. Le attuali iniziative di NBN in Germania¹⁰ come in Olanda, ad esempio, pur mantenendo una certa autonomia nell'implementazione del servizio (da qui il tentativo di avere policy e un sistema di risoluzione comuni a livello europeo sperimentato dal progetto PersID¹¹), hanno una chiara vocazione verso la digital preservation delle risorse identificate. Questa relazione è forte a tal punto che il riferimento alla conservazione delle risorse digitali identificate dal namespace NBN sarà esplicitato nella prossima versione dell'RFC 3188. Nel solco di questa visione si colloca l'attività della Fondazione Rinascimento Digitale nell'ambito del gruppo di lavoro IETF¹²

¹⁰http://www.persistent-identifier.de/english/3352-EPICUR_policy.php.

¹¹PersID- Building a persistent identifier infrastructure, <http://www.persid.org>. Il progetto PersID, finanziato dalla SURF Foundation, aveva come scopo la progettazione di un metaresolver tra i vari sistemi NBN europei.

¹²L'IETF ha attivato un gruppo di lavoro aperto per la revisione dello standard 3188.

per la revisione dell’RFC 3188 e di gruppi di progetto internazionali come PersID e la collaborazione con il Ministero dei beni Culturali (MiBAC) e la Conferenza dei Rettori delle Università Italiane (CRUI) per la realizzazione di un sistema analogo nel nostro paese.

Il servizio NBN Italia

Sulla base dei risultati raggiunti dalle prime ricerche sull’argomento, l’attuale gruppo di lavoro ha definito a livello generale l’organizzazione del sistema e le principali funzionalità. Ovviamente, l’iniziativa potrà prevedere evoluzioni e ottimizzazioni sulla base dei requisiti della comunità, che possono variare a partire dai livelli di performance del sistema fino alla distribuzione delle responsabilità dell’assegnazione dei nomi, come già ipotizzato (Bellini et al., «Persistent Identifiers distributed system for Cultural Heritage digital objects»; «Semantics-Aware Resolution of Multi-part Persistent Identifiers») e in parte sperimentato («Persistent Identifier Distributed System for Digital Libraries, Information Technology»; «Persistent identifiers: jNBN, a JEE application for the management of a national NBN infrastructure»). In questa fase, considerato il contesto di start up, si è privilegiato un’organizzazione più centralizzata per esigenze di controllo e assessment del sistema.

Il livello di persistenza

NBN si propone un livello di persistenza molto alto. Nel contesto URN si possono identificare almeno tre livelli di persistenza:

1. persistenza dell’identificatore. Se la risorsa non è più disponibile in rete, l’identificatore URN può continuare ad avere senso (ad esempio come prova che in un determinato momento quella risorsa esisteva);

2. persistenza dell'abbinamento URN e URL. Si tratta di un impegno che garantisce nel lungo periodo che un URN sia sempre risolto (porti almeno a un indirizzo di tipo URL). Non viene garantita l'accessibilità della risorsa ma viene garantito – in caso di non disponibilità permanente della risorsa in rete – l'accesso alla c.d. tombstone (es. “questo ebook non è più in commercio”);
3. persistenza della Risorsa referenziata da NBN. Assicurare nel lungo periodo l'esistenza e accessibilità della risorsa referenziata da URN. Questo è il livello di persistenza di NBN reso possibile solo dal deposito (legale o volontario) presso le biblioteche nazionali e dalla descrizione autorevole della Bibliografia Nazionale.

L'organizzazione

Il servizio NBN propone un'organizzazione su due livelli:

- il Registro Centrale dei nomi (RC) gestito dalla BNCF
- i Content Provider (CP), cioè coloro che hanno le risorse che devono essere inviate verso il deposito legale.

La richiesta di assegnazione del NBN avviene in modo decentrato direttamente dai CP autorizzati, seguendo un certo criterio di selezione delle risorse e di accreditamento dei CP stessi. Le istituzioni accreditate riceveranno un sub-namespace (in genere l'acronimo dell'istituzione) che verrà utilizzato per la generazione dei rispettivi NBN. Ad esempio l'Università di Firenze potrebbe ricevere il sub-namespace UNIFI. Il NBN risultante assumerà la forma: URN:NBN:IT:UNIFI-123456. Le istituzioni che riceveranno il sub-namespace dovranno inoltre necessariamente aderire a Magazzini

Digitali (Vignocchi, Bergamin e Messuti; «Le linee guida per l’accesso aperto alle tesi di dottorato») per il deposito legale dei contenuti da loro prodotti. Questo è evidentemente un processo off-line che deve prevedere la necessaria flessibilità per gestire casi specifici. Ad esempio, nel caso in cui una istituzione abbia digitalizzato dei contenuti, ma non sia in possesso di un IR digitale, le risorse potrebbero essere inviate alla BNCF tramite canale offline (es. CD, DVD, etc.).

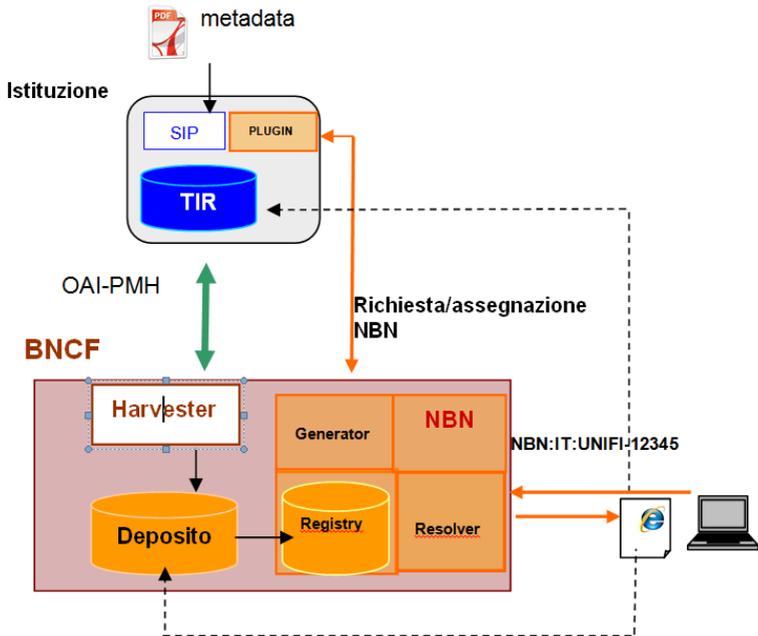


Figura 1: Schema descrittivo di NBN in relazione con il sistema di Deposito Legale

Come si evince dalla figura 1, il sistema NBN è formato da 3 macro blocchi funzionali lato centrale e un plugin lato CP. I tre blocchi

funzionali del sistema centrale sono: il Generatore, il Registro Centrale (RC) e il Resolver. Nei paragrafi seguenti vengono descritti gli elementi dal punto di vista delle loro funzioni principali tralasciando le modalità tecnologiche di implementazione.

Il Registro Centrale NBN

Il Registro Centrale (RC) è il cuore del sistema, il luogo dove vengono mantenute le associazioni NBN-URL. Una delle principali assunzioni di base per il RC è quella di non prevedere la gestione dei metadati descrittivi. Questa decisione si basa sia sulla complessità che sul costo di mantenere aggiornati i metadati presenti nei vari archivi collegati al deposito legale. In questo modo si evita di ricreare un nuovo catalogo e ci si focalizza sul servizio di identificazione, mantenendo invece un puntatore ai metadati autorevoli presenti negli archivi delle istituzioni. La distinzione tra fare un catalogo e un registro è sostanziale. Nel caso del RC di NBN si è provveduto a definire una serie di informazioni gestionali del nome stesso per monitorarne lo stato, la tracciabilità delle modifiche, etc. Il registro quindi mantiene tutte le richieste, le azioni, le modifiche che avvengono sul registro stesso al fine di garantire il trust complessiva del servizio.

Il Generatore NBN

Un NBN viene assegnato a quello che nella terminologia PREMIS¹³ viene identificato come *representation*. In altre parole un record formato da più file riceve un solo NBN. Inizialmente il NBN sarà assegnato alle tesi di dottorato visto che questa tipologia di oggetti

¹³Preservation Metadata: Implementation Strategies (PREMIS) è lo standard di metadati per la conservazione promosso da OCLC e RLG, <http://www.loc.gov/standards/premis>.

è già oggetto di harvesting e deposito da parte di Magazzini Digitali, ma è pensato per essere assegnato a tutte le risorse destinate al deposito legale e alle risorse frutto dei progetti di digitalizzazione condotti dalle biblioteche e da altre istituzioni. La stringa alfanumerica è opaca e non darà alcuna informazione semantica (es. uso di parole specifiche), inoltre, avrà una lunghezza consona alle esigenze di citazione. Si è voluto evitare la complessità d'uso, ad esempio, dell'Universally Unique IDentifiers (UUID) che è problematico a causa della sua lunghezza di 36 caratteri complessivi¹⁴. Pertanto si è ritenuto opportuno generare un progressivo generale che non sia legato al sub-namespaces, ma si basi su un contatore unico.

Il Resolver NBN

Il Resolver¹⁵ è lo strumento che consente l'accesso alla risorsa ma allo stesso tempo anche la sua verifica, che se positiva, consolida e aumenta il trust del sistema, come precedentemente spiegato. Il Resolver è quindi l'interfaccia del servizio verso l'utente e il tipo di risposta che dà è effettivamente il livello di servizio offerto dal sistema. In generale una risoluzione di un NBN prevede la restituzione di una pagina intermedia con

1. i metadati e il link alla risorsa presente nel IR del content provider ritenuto Trusted,
2. le informazioni di stato del NBN e
3. il riferimento alla copia della risorsa presente nel deposito legale.

Questo è il livello di servizio principale dal quale poi è possibile evolvere verso sistemi più sofisticati in accordo alle esigenze della

¹⁴UUID, http://en.wikipedia.org/wiki/Universally_unique_identifier.

¹⁵Il dominio di risoluzione sarà <http://nbn.depositolegal.it>.

comunità. È importante notare che il servizio di deposito legale ha tempi di harvesting diversi rispetto al flusso di pubblicazione delle risorse. L'intero workflow include quindi la presenza di un periodo di latenza tra la richiesta di registrazione del NBN per la risorsa e la sua effettiva archiviazione nel sistema MD. Questa latenza, però non andrà a inficiare il livello di servizio di risoluzione offerto. Infatti si assume che il servizio di risoluzione centrale sia in grado di risolvere il NBN dal momento in cui viene generato, a fronte della richiesta da parte dei CP e non solo dopo che è stato archiviato.¹⁶

Il plugin per gli IR

Uno degli aspetti più critici di un sistema di identificazione non è la generazione del nome o il processo al termine del quale l'identificativo viene assegnato, ma la sua concreta consegna al richiedente. La consegna implica una sorta di "uscita" dell'identificativo dal sistema generante per andare a inserirsi in uno spazio codificato del richiedente (il posto della targa, la carta di identità, la tesserina del codice fiscale, etc.). A questo proposito l'obiettivo più importante era la consegna contestuale alla richiesta (quindi immediata) del NBN generato, così da entrare subito a far parte dei metadati descrittivi della risorsa. Questo approccio si differenzia, migliorandolo, da quello fino a oggi adottato dalla Biblioteca nazionale tedesca che prevede una attesa di 24/48 ore. Per ottenere questo risultato importante sia in termini di usabilità che di efficienza del sistema è stato sviluppato un apposito plugin, al momento adattabile per il software EPrints¹⁷ che consentirà il dialogo con il sistema centrale tramite web service.

¹⁶L'interfaccia progettata per essere multilingue presenterà le informazioni necessarie sia sul servizio sia sulle modalità di partecipazione.

¹⁷È previsto a breve anche lo sviluppo di un plugin per DSpace, che con EPrints costituisce il software più utilizzato dalle università italiane per l'implementazione di open archives.

Il plugin modifica l'interfaccia del sistema di archiviazione usato dall'IR inserendo un elemento attraverso il quale l'operatore sarà in grado di richiedere la generazione del NBN da remoto. Questo strumento, come si vede dalla figura 1 a pagina 15, si inserisce nel workflow di pubblicazione e consente all'operatore di effettuare una richiesta di generazione del NBN prima della validazione finale della risorsa (pubblicazione) e solo per quelle risorse che già si intende inviare al deposito legale.

Sperimentazione e nuove prospettive

Grazie allo sviluppo dei plugin sarà possibile avviare una prima sperimentazione del sistema NBN in collaborazione con il gruppo di lavoro Open Access della Commissione biblioteche della CRUI, coordinato da Roberto Delle Donne. Il gruppo di lavoro ha collaborato con il MiBAC e in particolare con le Biblioteche Nazionali Centrali alla implementazione di una procedura automatizzata per il deposito legale delle tesi di dottorato di ricerca¹⁸ - i primi contenuti digitali a popolare Magazzini Digitali (Vignocchi, Bergamin e Messuti; «Le linee guida per l'accesso aperto alle tesi di dottorato»). Saranno proprio le tesi di dottorato di ricerca di alcune università che hanno già avviato con successo il deposito legale in Magazzini Digitali a ricevere i primi NBN e a testare le funzionalità del sistema.¹⁹ Lo sviluppo di NBN beneficerà anche di un "ecosistema" costituito da una serie di attività correlate con i PI a cui che i partner dell'iniziativa partecipano. Ad esempio, la Fondazione Rinascimento Digitale è leader del workpackage sulla citabilità e gli identificatori nel progetto europeo APARSEN. Tra i risultati del progetto vi è il

¹⁸<http://www.cruil.it/homepage.aspx?ref=1891>

¹⁹La CRUI ha incaricato di questa prima sperimentazione l'Alma Mater Studio-rum, Università di Bologna e l'Università di Trieste rispettivamente per testare la compatibilità di archivi EPrints e DSpace.

rilascio di un nuovo framework di interoperabilità tra i vari sistemi di identificazione non solo per gli oggetti (NBN, DOI, Handle, etc.) ma anche per gli autori e le istituzioni. L'intento del documento è quello di definire una piattaforma comune di interoperabilità in grado di mettere in relazione queste entità e creare i presupposti per lo sviluppo di servizi avanzati a carattere sia metrico che semantico. Ai lavori partecipano organizzazioni come il CERN, l'Università di Trento, le biblioteche nazionali tedesca e inglese, gli archivi inglesi e molti altri partner. Le attività di sviluppo di NBN stanno già tenendo in considerazione questa prospettiva e i nuovi sviluppi saranno intesi a rendere il sistema, primo in Italia, conforme al framework. Infatti il sistema NBN risponde ai criteri di trust identificati all'interno di APARSEN per i sistemi di PI. I principali criteri a cui NBN corrisponde positivamente sono:

- esistenza di una Registration Authority;
- esistenza di un Resolver accessibile su Internet;
- garanzia dell'unicità dell'identificativo dentro il contesto (dominio) di riferimento;
- garanzia della persistenza del PI (sia a livello di funzionalità che di business model);
- esistenza di policy per gli archivi che ricevono il PI per le proprie risorse;
- risoluzione affidabile;
- disaccoppiamento tra la specifica del PI e la URL di risoluzione.

In particolare, nel confronto con gli attuali sistemi, NBN:IT è del tutto paragonabile al livello di trust ottenuto dal sistema della biblioteca nazionale tedesca sui medesimi criteri.

Un'altra importante iniziativa assunta dalla CRUI partner del consorzio NBN, è la responsabilità della gestione del cosiddetto DOI non commerciale. La CRUI, stipulando un accordo con DataCite, si è attivata per la distribuzione del DOI in particolare alle istituzioni scientifiche che vogliono identificare dataset, risultati della ricerca, articoli, etc. La simultanea presenza nel gruppo di NBN dei responsabili diretti di queste iniziative, consentirà straordinarie sinergie e uno sviluppo coerente di questi servizi con notevoli benefici per l'accesso, il riuso e la diffusione della conoscenza.

Conclusioni

La realizzazione di un sistema di identificazione persistente legato alla digital preservation si propone come parte integrante delle infrastrutture nazionali e internazionali per la ricerca e i beni culturali in primis. L'adozione di identificativi persistenti contribuisce alla definizione dei trust necessaria in particolare per quelle risorse sulle quali la società costruisce e stratifica il nuovo sapere. Occorre dire naturalmente che NBN non intende essere una soluzione alternativa agli identificatori in uso nel mondo dei Linked Data. NBN è un identificatore indipendente dal protocollo di risoluzione ma può essere usato all'interno di Cool URI e offrire un importante servizio anche per i Linked Data (ad esempio il NBN URN:NBN:IT:UNIFI-123456 può essere usato come Cool URI nella forma <http://nbn.depositolegale.it/URN:NBN:IT:UNIFI-123456>).

Ovviamente, come ogni servizio, il suo successo sarà valutato dal livello di integrazione negli attuali workflow di pubblicazione e produzione di contenuti, dall'accettazione da parte delle comunità e dalla sostenibilità economica. Per favorirne il successo, il consorzio si impegna a tenere in considerazione il contesto internazionale, rece-

pendone le principali direzioni e raccomandazioni, pur adattandole al contesto specifico italiano.

Riferimenti bibliografici

- Bazzanella, Barbara, et al. «Persistent Identifiers Interoperability Framework». (2012). (Cit. alle pp. 2, 3).
- Bellini, Emanuele e Paolo Ceravolo. «EPICA: Easy Persistent Identifier Common Architecture». *Lecture Notes in Computer Science* 6428. (2010): 626–634. (Cit. a p. 6).
- Bellini, Emanuele e Maurizio Lunghi. «Persistent Identifiers for cultural heritage». (2007). <http://www.digitalpreservationeurope.eu/publications/briefs/persistent_identifiers.pdf>.
- Bellini, Emanuele, et al. «Persistent Identifier Distributed System for Digital Libraries, Information Technology». **proceedings** di World Library And Information Congress: 75th Ifla General Conference And Council, Milan. 2009. (Cit. alle pp. 6, 13).
- Bellini, Emanuele, et al. «Persistent Identifiers distributed system for Cultural Heritage digital objects». *iPRESS 2008 Proceedings*. 2008. 55–65. (Cit. alle pp. 5, 6, 13).
- Bellini, Emanuele, et al. «Persistent identifiers: jNBN, a JEE application for the management of a national NBN infrastructure». *GARR Conference 2009*. Rome: GARR, 2009. (Cit. alle pp. 6, 13).
- Bellini, Emanuele, et al. «Semantics-Aware Resolution of Multi-part Persistent Identifiers». *Emerging Technologies and Information Systems for the Knowledge Society*. A cura di Miltiadis Lytras, et al. Vol. 5288. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 2008. 413–422. <http://dx.doi.org/10.1007/978-3-540-87781-3_45>. (Cit. alle pp. 5, 6, 13).
- Berners-Lee, Tim. «Cool URIs don't change». *W3C Design Issues*. (1998). <<http://www.w3.org/Provider/Style/URL.html>>. (Cit. a p. 3).
- «CENL Task Force on Persistent Identifiers, Report 2007». (2007). <<http://pid.ndk.cz/dokumenty/dalsi-cetba/cenl-task-force-on-persistent-identifiers-report-2007>>. (Cit. alle pp. 6, 7).
- Daigle, Leslie L., et al. «URN Namespace Definition Mechanisms - RFC 2611». (1999). <<https://ebook.tools.ietf.org/html/rfc2611>>. (Cit. a p. 11).
- «Final Report of the Project Team of CEN/ISSS eGovernment Focus Group, CEN». (2008). <<http://www.egovpt.org/fg/Report>>. (Cit. a p. 8).
- Hakala, Juha. «National Bibliography Numbers as Uniform Resource Names – NBN RFC 3188». (2001). <<http://tools.ietf.org/html/rfc3188>>. (Cit. alle pp. 11, 12).
- «ISO 3297:1986 Documentation - International standard serial numbering (ISSN)». (1986). <http://www.iso.org/iso/iso_catalogue/catalogue_ics/catalogue_detail_ics.htm?csnumber=8534>. (Cit. a p. 11).

- «Le linee guida per l'accesso aperto alle tesi di dottorato». *AIDAinformazioni* 26.3-4. (2008): 73–89. <<http://hdl.handle.net/10760/12459>>. (Cit. alle pp. 6, 15, 19).
- Luhmann, Niklas. «Trust: a mechanism for the reduction of social complexity». *Trust and Power*. A cura di Niklas Luhmann. New York: Wiley, 1979. 4–103. (Cit. a p. 8).
- Lynch, Clifford, Cecilia Preston e Ron Jr. Daniel. «Using Existing Bibliographic Identifiers as Uniform Resource Names, RFC 2288». (1998). <<http://www.ietf.org/rfc/rfc2288.txt>>. (Cit. a p. 11).
- Masinter, Larry e Karen Sollins. «Functional Requirements for Uniform Resource Names, RFC 1737». (1994). <<http://www.ietf.org/rfc/rfc1737.txt>>. (Cit. a p. 11).
- Moats, Ryan. «URN Syntax, RFC 2141». (1997). <<http://tools.ietf.org/html/rfc2141>>. (Cit. alle pp. 3, 11).
- «NESTOR Catalogue of criteria for assessing the trustworthiness of PI systems». (2009). <http://files.d-nb.de/nestor/materialien/nestor_mat_13_en.pdf>. (Cit. a p. 2).
- «NISO/ANSI Z39.56-1997 Serial Item and Contribution Identifier». (1997). (Cit. a p. 11).
- «NISO/ANSI/ISO 2108:1992 Information and documentation - International standard book number (ISBN)». (1992). (Cit. a p. 11).
- Solimine, Giovanni. *Controllo bibliografico universale*. Milano: Bibliografica, 1998. (Cit. a p. 11).
- Sollins, Karen. «Principles of Uniform Resource Name Resolution IETF, RFC 2276». (1998). <<http://www.ietf.org/rfc/rfc2276.txt>>.
- «The National Libraries Resolver Discovery Service (RDS) - CENL Recommendation». (2007). <<http://pid.ndk.cz/dokumenty/dalsi-cetba/the-national-libraries-resolver-discovery-service-a-cenl-recommendation>>.
- Vignocchi, Marialaura, Giovanni Bergamin e Raffaele Messuti. «Tesi di dottorato: stato dell'arte, iniziative in corso, scenari possibili». *Bibliotime* 13.3. (2010). <<http://didattica.spbo.unibo.it/bibliotime/num-xiii-3/vignocchi.htm>>. (Cit. alle pp. 6, 15, 19).
- Wilson, Frank. «Trust and Identity in Interactive Services: Technical and Societal Challenges». *cc:eGov Organisational change for citizen-centric eGovernment – Think Paper* 11. (2011). <<http://zealot.mrnet.pt/govnet/images/articles/5019/PaperTrustandidentityininteractive.pdf>>. (Cit. a p. 8).
- Winer, Dov. «Persistent Identifiers systems in the Public Administration sector». (2008). <http://www.digitalpreservationeurope.eu/publications/briefs/persisten_identifiers_for_public_administration.pdf>. (Cit. a p. 8).

EMANUELE BELLINI, Fondazione Rinascimento Digitale
bellini@rinascimento-digitale.it

CHIARA CIRINNÀ, Fondazione Rinascimento Digitale
cirinna@rinascimento-digitale.it

MAURIZIO LUNGI, Fond. Rinascimento Digitale
lunghi@rinascimento-digitale.it

CINZIA LUDDI, Fond. Rinascimento Digitale
cinzia.luddi@gmail.com

MAURIZIO MESSINA, Biblioteca Nazionale Marciana
messina@marciana.venezia.sbn.it

GIOVANNI BERGAMIN, Biblioteca Nazionale Centrale di Firenze
giovanni.bergamin@bncf.firenze.sbn.it

RAFFAELE MESSUTI, Biblioteca Nazionale Centrale di Firenze
raffaele@atomotic.com

GIOVANNA CORDANI, Biblioteca Nazionale Centrale di Roma
g.cordani@bnc.roma.sbn.it

ROBERTO DELLE DONNE, Conferenza dei Rettori delle Università Italiane
delledon@unina.it

FRANCESCA ROSSI, Conferenza dei Rettori della Università Italiane
rossi@cru.it

MARIALAURA VIGNOCCHI, Università di Bologna
marialaura.vignocchi@unibo.it

STEFANIA ARABITO, Università degli Studi di Trieste
arabito@units.it

JORDAN PIŠČANC', Università degli Studi di Trieste
piscanc@units.it

Bellini, E., C. Cirinnà, M. Lunghi, et al. "Il progetto National Bibliography Number Italia (NBN:IT).

Un identificatore persistente a supporto del deposito legale nazionale delle risorse

digitali". *JLIS.it* Vol.3, n.1 (Giugno/June 2012): 4789-1-4789-26. DOI: [10.4403/jlis.it-4789](https://doi.org/10.4403/jlis.it-4789). Web.

ABSTRACT: L'associazione di un identificatore persistente a una risorsa digitale è considerata oggi una best practice che contribuisce alla risoluzione del problema del reperimento affidabile delle risorse in rete. Attualmente si stanno affermando diverse tecnologie e standard come DOI, Handle e Cool URI con differenti peculiarità e livelli di servizio offerti. Tuttavia, all'interno delle comunità scientifiche e culturali e nel nascente semantic web, persiste la necessità di garantire l'autenticità, la provenienza e in particolare la conservazione delle risorse nel tempo. Il National Bibliography Number (NBN) nasce nel contesto del ruolo e delle responsabilità che le biblioteche nazionali hanno nel campo del controllo bibliografico universale ed è orientato verso l'identificazione di risorse che devono essere conservate nel lungo periodo. Il progetto coordinato dalle biblioteche nazionali di Firenze, Roma e Venezia, dalla Fondazione Rinascimento Digitale e dalla CRUI non è alternativo ai sistemi di PI attualmente in uso, ma ne completa l'offerta.

KEYWORDS: Conservazione digitale; Deposito legale; Identificatori persistenti; National Bibliographic Number; Web semantico

Submission: 2012-02-13
Accettazione: 2012-03-21
Pubblicazione: 2012-06-01





Italian SSH journals in Journal Citation Reports (JCR) and in SCImago Journal Rank (SJR): data and first analysis

Andrea Capaccioni, Giovanna Spina

Introduction

Upon the publication of the announcement for the participation (7th November 2011) in the Framework for Research Quality Assessment 2004-2010 (VQR 2004-2010), published by Agenzia di Valutazione del Sistema Universitario e della Ricerca (ANVUR), it began to operate a complex organisational process which aim is to assess a representative sample of the scientific production of Italian Universities within the last 10 years.¹ The importance of research assessment exercise (RAE according to the English definition) in the area of scientific research products (articles, books, patents, etc.) has increased consistently in the first decade of the new century (an introduction

¹The organisation and the management of the assessment framework are detailed in the founding decree of ANVUR (DPR n. 76 of the February the 1st 2010) and in DM of July the 12th (<http://www.anvur.org>). The text announcing the participation call is available at: http://www.anvur.org/sites/anvur-miur/files/bando_vqr_def_07_11.pdf.



to the topic can be found in Baccini, p. 11-35; De Robbio). Recently, in addition to the Research Quality Assessment 2004-2010 launched in Italy, the English Research Excellence Framework (REF) 2009-2014² and the Excellence in Research for Australia initiative (ERA)³ launched in 2010 have to be mentioned. The governments became ever more careful in establishing precise criteria for awarding funds for research, thus making the scientific institutions more functional, stimulating a greater efficiency of research, allocating resources on the basis of merit, and demonstrating that investments produce a common good (Abramo, D'Angelo, and Di Costa, p. 929-941). The most debated issue is, of course, what the evaluation criteria are. Unique criteria or different criteria for different areas? Qualitative criteria (peer review) or quantitative criteria (Impact Factor, etc)? Bibliography concerning this topic increased exponentially in the last decades (for a historical reconstruction of bibliometrics and current debate see De Bellis). In the following pages we mean to provide a paper analyzing the space reserved to Italian journals in the field of Humanities and Social Sciences within the main bibliographic tools used for the evaluation of research (for more information on the topic see Graziosi, p. 31). We will concentrate on bibliographic databases such as Web of Science and Scopus.⁴ We believe it is important to deepen the knowledge on these topics that is not currently developed, through a basic presentation of the main aspects of the issue. Data on Italian journals, drawn from the chosen databases, are going to be entirely presented and subjected to critical examination. Web of Science (WoS), created by Thomson Reuters (already

²<http://www.hefce.ac.uk/research/ref/>.

³<http://www.arc.gov.au/era/>.

⁴ANVUR explicitly indicates the two databases in its document "Criteri e parametri di valutazione dei candidati e dei commissari dell'abilitazione scientifica nazionale", 22nd June 2011, http://www.anvur.org/sites/anvur-miur/files/documento01_11.pdf

part of the Institute for Scientific Information ISI), is an online bibliographic commercial database (it contains full text articles and abstracts) accessible through the Web of Knowledge portal,⁵ and it is articulated in three sections (*Science Citation Index*, *Social Science Citation Index*, *Arts & Humanities Citation Index*). WoS aims to select journals from different disciplines with rigorous criteria.⁶ Related to WoS is the *Journal Citations Reports* (JCR) which gathers more than 10,000 journals belonging to more than 230 disciplinary areas, published by more than 2,000 publishers. JCR is used as a tool to compare and evaluate journals. JCR uses bibliometric indicators, some traditional like Impact Factor (IF) and Five-year IF, to others that are considered new metrics such as the Eigenfactor, etc. The annual edition contains data related to the publications of the previous year. JCR is available in two editions: *JCR Science Edition* covering more than 8,073 journals, and *JCR Social Science Edition* with 2,731 journals.⁷ The other data source considered for this study is Scopus, an initiative launched by Elsevier in 2004.⁸ Scopus is a bibliographic commercial list that - compared to WoS - is more extensive when it deals with languages and countries of publication of the journals scope of this study. Furthermore, it contains a larger number of journals belonging to the SSH (Social Sciences and the Humanities) area (see table 1).⁹ Our analyses are based on the free access database Scimago, created in collaboration between some Spanish universities and the publisher Elsevier in 2007. The bibliometric indicator developed by Scimago is the Journal Rank Indicator (SJR), whose

⁵<http://wokinfo.com/125>

⁶http://wokinfo.com/products_tools/multidisciplinary/webofscience/.

⁷http://thomsonreuters.com/products_services/science/science_products/a-z/journal_citation_reports/.

⁸<http://www.scopus.com>.

⁹Tables are available as a separate file at <http://leo.cilea.it/index.php/jlis/article/downloadSuppFile/4787/180>.

sources are the journals indexed by Scopus starting from 1996. SJR is the main competitor of IF used by the Journal Citation Report (WoS). SJR is developed with the algorithm Google PageRank and, besides calculating the impact by journal, it calculates the impact by Country using the function Country Rank. SJR was developed by Félix de Moya from the Consejo Superior de Investigaciones Científicas (Spain) and by Vicente Guerrero Bote from the Universidad de Extremadura (Spain) (for further readings see Gonzalez-Pereira, Guerrero-Bote, and Moya-Anegón, p. 379-391). The results of the analyses are available free of charge on the SCImago Website.¹⁰

Methodology

This paper examines the journals (excluding books, reports, etc.), part of the two above-mentioned indexes, responding to the following characteristics. They have: a) to be published in Italy (not exclusively in Italian language); b) to belong to SSH area. For a better identification of the journals in the list we referred to the ISSN of the review. Data are taken from three sources: a) the *Journal Citation Reports* (JCR) of WoS with reference to *JCR Social Sciences edition* for the year 2010. The examination of JCR was carried out using the online access made available by the Library Services Center of the Università degli Studi di Perugia; b) the *Arts & Humanities Citation Index* (A&HCI) of WoS for the year 2010. WoS does not calculate the IF of the journals. The examination of this bibliographic database will serve to enrich the information on the consistency of the presence of Italian journals in WoS, provided that it cannot be used for evaluation purposes. The examination of the A&HCI was carried out using the online access made available by the Library Service Center of the Università degli Studi di Perugia; c) SCImago Journal

¹⁰<http://www.scimagojr.com/>.

& Country Rank (SJR). Searches on SJR was restricted to subject areas: *Arts and Humanities and Social Sciences* for the year 2010. SJR shows bibliometric indicators even for Humanities journals. The last examination of SJR was made through the portal SCImago Journal & Country Rank in January 2012. We took into account also the information available on the official websites of the two bibliographic databases and the most important bibliographic papers on the topic. For the differences between JCR and SJR we take as a reference a selection from the vast literature on the subject (there are many papers suggesting a comparative analysis between the two data bases, among which we indicate: Deis and Goodman, p. 5-21; Jacsò, "ISI Web of Science, Scopus, and SPORTDiscus", p. 51-54; p. 1537-1547; La Guardia; Tarantino, p. 23-32; Anegòn et al., p. 53-78).

The data

In this section we will present the collected data and their critical analysis. We consider appropriate to show in full the data obtained from the research on the bibliographic databases. The presentation of data should be considered as an integral part of the paper and aims to promote a better understanding of the proposed analysis, in order to help understanding the functioning of the databases and the bibliometric indicators which are being used, and finally to allow those who are interested (scientists, evaluators, librarians, etc.) to study in depth the issue. Data are summarized in tables 2 to 12. Table 2 shows, in alphabetic order, the Italian journals present in JCR Social Science Edition 2010.¹¹ The Italian journals in the section Arts

¹¹JCR is available in two editions: JCR Science Edition and JCR Social Science Edition. Each annual edition contains data related to the publications of the previous year, and shows connection between the journals with citations and citing journals in a simple and easy to use way

& Humanities Citation Index 2010 are listed in table 3, in alphabetic order as well. Table 4 shows Italian journals in SJR Social Sciences 2010 and table 5 contains Italian journals in SJR Arts & Humanities 2010. The three following tables show Italian position (in terms of number of publications) in the world rankings (2010) calculated by SJR, respectively in the area of Humanities (table 6), in the area of Social Sciences (table 7), and in all disciplines (table 8). Table 9 outlines the position of Italy in terms of number of publications, as listed in WoS (data are related to the year 2010). The Venn's diagram (fig. 1 on page 9) represents the total number of Italian journals belonging to all the disciplines listed in Scopus and WoS. The last two tables (10, 11) compare Italian journals of the Humanities area and those of Social Sciences present both in JCR and SJR (2010). Scopus includes not only a large collection of journals (18.854 compared to 10.804) but also a great number of represented countries (231 compared to 84) and a great (50 compared to 45) language variety (see table 1). The journals published in languages other than English are actually underrepresented in both databases. Chronological coverage of Scopus is less extensive than the one of WoS. Concerning to bibliometric indexes, the two databases use different tools. SJR indicator, for instance, assumes that not all the citations are equal, therefore it introduces a careful consideration based on the prestige of the citing journal and excludes self-citation. In the 2010 JCR Social Science edition, 13 Italian journals were indexed: they represent 0.47% of the total number of the listed reviews (2,731). The percentage of the journals published in the United States and England, indexed by JCR, is respectively 45% and 26.3%. Therefore the journals of those geographical areas represent 71.3% of the total. Furthermore the presence of Italian journals turns out to be reduced compared to other European nations. For instance, there are 52 Spanish journals which equals to 1.90% of the total number of journals in JCR Social

Sciences, and 25 French journals corresponding to 0.91%. We should keep in mind that JCR adopts its own criterion of disciplinary classification which includes medical journals among Social Sciences publications (table 2). We are going to take into account now the Impact Factor (IF) and the *five-year* IF of some Italian journals which are present in the table 2.¹² The Impact Factor of the Journal of Anthropological Sciences of the Italian Anthropological Institute (Istituto Italiano di Antropologia) equals to 2.000 and guarantees to the journal the seventh place (of 76) among the journals belonging to this area and indexed by JCR. The journal with the highest IF in this category is Evolutionary Anthropology with the value 4.531. Cadmo, Italian journal of experimental pedagogy, is positioned 166 (of 184) in the category Education & Educational research with IF equal to 0.160. The journal with the highest IF in the same category is Educational Research (USA) with IF 3.774. Economia Politica records an IF equal to 0.610, not a very high level in the area of Economics. The Journal of Economic Literature (USA) is positioned first with an IF equal to 7.432. Nuncius. Journal of the History of Science (category of History and Philosophy of Science) has an IF value of 0.080, while the highest value (3.986) in this category was attributed to the US journal American Journal of Bioethics. The journal belonging to the Social Science Edition 2010 with the absolute highest IF is the English Behavioural and Brain Sciences (21.952). We underline that History and Philosophy of the Life sciences, edited by Stazione Zoologica Anton Dohrn from Naples, published by an Italian pub-

¹²The authors remind readers that annual IF (in this case the 2010) is calculated by dividing the number of citations received by journals in 2010 on the articles published in 2008 and 2009 by the overall number of articles published during those two years. The *five-year* IF (for a five-year period) corresponds instead to the average number of times the articles of the journals published in the last five years were cited in JCR year, in our case 2010. This measuring parameter, available only from JCR 2007, is used to better analyse the impact of the journals in the areas where the influence of public research is developed for a greater period of time.

lisher (F. Giannini, Napoli), was not included among the thirteen Italian journals, but among the English journals. This inconsistency, although rare, reveals the need for an examination of the use of classification criteria adopted by databases. *Arts & Humanities Citation Index* (A&HCI) of WoS analyses the humanities related publications (with abstracts and bibliographic information), but does not offer instruments for bibliometric measurement. A&HCI 2010 indexed 1679 journals out of which 63 Italian equal to 3.75% of the total (table 3). We underline a considerable percentage rise of the relevance of Italian journals in A&HCI compared to JCR Social Sciences. During 2010, SJR Social Sciences (Scopus) indexed 2,958 journals in the social sciences area, out of which 1.14% (34) are Italian (table 4); while 1,638 are journals belonging to the area of humanities out of which 55 are Italian, equal to 3.35% (table 5). The total of Italian journals (and other kinds of material) in the database is 323, that is 1.71% of the total number of indexed documents (18,854). It should be noted that, unlike WoS, Scopus began indexing also monographs and series as a recent insertion in its database of the titles belonging to the European Reference Index for Humanities (ERIH) of European Science Foundation shows.¹³ The tables 4 and 5 show the list of Italian journals included in humanistic and social sciences area of SJR Social Sciences even in order to facilitate their identification. The title of a single journal was flanked by ISSN and the values of the following indicators: total cites,¹⁴ SJR, h-Index,¹⁵ In the 2010

¹³Scopus works with European Science Foundation to expand Arts and Humanities coverage, http://www.elsevier.com/wps/find/authored_newsitem.cws_home/companynews05_01241.

¹⁴Scimago total cites is an indicator which gathers the total of citations received in other articles published in a journal in the last three years.

¹⁵H-index is an index elaborated by Jorge E. Hirsch (University of California) in order to measure both productivity and impact of the work of the scientists, on the basis of the number of their publications and the number of citations they have received in other publications. Hirsch's definition is as follows: a scholar has an h-

ranking drawn up by SJR, Italy holds eighth position (of 146 countries) regarding publication of documents (articles etc.) belonging to the area of humanities, preceded by the United States, United Kingdom, France, Canada, Germany, Australia and Spain¹⁶ (table 6). In the social sciences area, our country takes eleventh position out of 195 (table 7). Regarding all the areas, it is positioned eighth among 231 countries (table 8). In order to make the comparison easier, we have shown in table 9 the level of Italian journals presence within WoS regarding the disciplinary areas which are subject matter of this paper. The following tables, as specified, have a comparative purpose. The diagram of Venn (fig. 1) represents the total number of Italian journals, which belong to all disciplinary areas, in the Scopus database (323) and in WoS (197). In reality, the total number of jour-

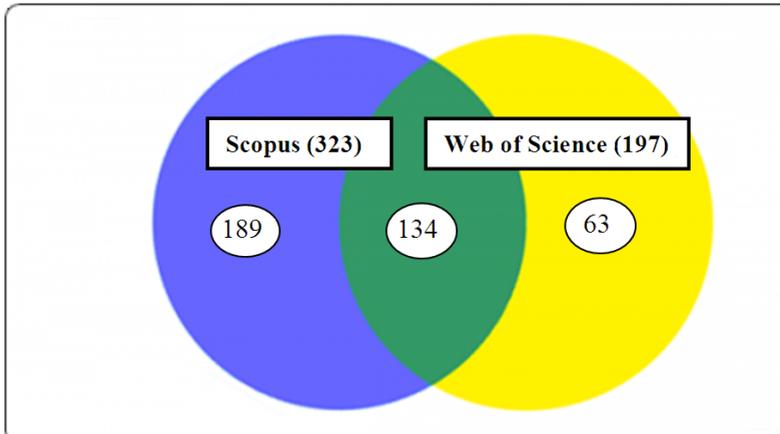


Figure 1: Italian journals present in Scopus and WoS

nals represented in the two databases, as the following data show,

index 50 if he has 50 articles with at least 50 citations each.

¹⁶www.scimagojr.com/countryrank.php

is 386. The results in detail: shared journals, (those present in both databases) are 134; the journals present exclusively in one database are respectively 189 in Scopus and 63 in WoS. These data confirm a greater coverage of the Italian production by Elsevier (Scopus). The table 10 compares the bibliometric values of the 13 journals that can be found in both databases, even if they are placed in different disciplinary areas. While in WoS all journals are classified in the area of Social Sciences, in Scopus the same journals are collocated in different areas: nine journals are placed in medical area, the remaining four are inserted into the Social Sciences one (Cadmio, *Economia Politica*, *International Journal of Transport Economics*, *Journal of Anthropological Sciences*). Finally, the table 11 shows the journals present both in Arts & Humanities Citation Index (bibliometric value not calculated by the producer) and in SJR Arts & Humanities 2010 (with bibliometric value).

Conclusion

The issue of the space reserved to the scientific production of Humanities and Social Sciences areas in bibliographic databases has two main aspects: a general level related to the bibliometric topics and to the evaluation policies, and a more specific one which deepens the adequacy of bibliographic instruments (databases, journal classifications and indexes, etc.) used at present in evaluating processes. In the first case, during the last few years there has been a criticism regarding the use of unique criteria of evaluation for both Hard sciences and Humanities and Social Sciences. In this way the claim of superiority of the former sciences over the latter ones actually fails, since the comparison is done with non adequate parameters. In a recent paper, Andrea Bonaccorsi (of the Executive Council of ANVUR) explained how the literature on the subject

has now become aware of a differentiated use of evaluating criteria (Bonaccorsi). It is high time for us to support the search for criteria capable of understanding with a greater coherence the values expressed by scientific production of Humanistic and Social area. An clear example is the debate on the typologies of scientific products, in particular on monographs. Bibliographic databases pay very little attention to monographs, and we could notice it even with the products considered for this article. The main databases involved in the evaluation activity tend to overestimate the role of the article, supporting instead those disciplinary areas, particularly ones belonging to hard sciences, that now have ceased to resort to the monographic studies. Paradoxically, Bonaccorsi continues, every bibliometric activity achieved with the present criteria would result "futile, because it would measure a marginal part of Humanities production (articles in journals), leaving completely unchanged the monographic production". During a recent seminar entitled University Research and its Evaluation organized by Università Cattolica (Milano, 28 settembre 2011), Edoardo Barbieri and Giovanni Solimine presented a decalogue on research evaluation in the area of Humanities (Barbieri). Talking about monographs, the two scholars pointed out that "for Humanities and Social Sciences, the scientific journal - on which impact indicators are traditionally based - is not the most popular type of publication, or however it is not the only one. Alongside papers on scientific journals, these disciplines commonly use the specialized monograph, which is the most common type of publication and which does not coincide with the production of manuals. Among the different types of monographs, some sectors privilege the editions of tests (accompanied by an adequate philological apparatus and/or a comment), the editing of inventories and catalogues, the records of conventions and symposia" (, p. 34). The evaluation activities of Humanities and Social disciplines shall

therefore take into account some specific elements, and particularly the extensive use of the monograph format. There are, as mentioned in the paper, positive signs of change in trends which give us hope. A cautious opening by the bibliographic databases producers towards monographs, especially Elsevier (Scopus), can be noted, too. The producers of databases follow the debate on the scientific production evaluation, and they have realised that the competition is getting keener. Until the beginning of the new century, WoS was almost the sole bibliographic tool used in evaluation processes, but with the introduction of Scopus (2004) this supremacy was reduced. We should take into account that during the last decade, National authorities in charge of research policy increased the demand for more reliable evaluation tools. The current crisis has made urgent a selection of worthy academic and research institutions in order to take the maximum advantage of an ever reduced funding. These are the main reasons that urged bibliographic databases producers to consider the enlargement of the number of analyzed products. The studies that intend to analyse the compliance of single databases with the requirements of evaluative activities are placed on a more specific level. From the data and the analyses exposed in the previous paragraph a substantial confirmation of a general condition of poor interest towards Humanities and Social Sciences emerges, as well as the presence of some less negative signals. In WoS, Humanities make up 15.2% of the entire database and in Scopus 8.68%. Nevertheless, we can note that Italian journals which belong to the area of Humanities and Social Sciences are in a satisfactory position. In SJR Social Sciences, Italian scientific production takes eighth place (table 6) in country ranking, confirming the same position that Italian scientific production holds as a whole (all the disciplines) within the same database (table 8). A better result is registered in WoS. Table 9 shows how Italian Humanities journals, again in

country ranking, gain a better position (sixth position) compared to the Social Sciences journals (eleventh position) and to the journals of all disciplinary areas (eleventh position). In a recent paper Emanuela Reale, revising the data of Erawatch-Metris, confirmed the conflicting signals regarding Italian scientific production in the area of Humanities and Social Sciences (Reale, p. 21-22). Reale emphasized the increase of the total number of Italian publications in the area of social sciences and humanities: from 32547 works in 2000 we moved to 43758 in 2006. The number of publications records an average increase of 7.2% (compared to 1.5% of France, 5.9% of Germany, 3% of Great Britain, only Spain has a better performance with 11%).¹⁷ The scholar also emphasized a satisfactory affirmation of the overall Italian production which for the 2006 reached 10% of the sector's total, against 19.4% of Great Britain, 18% of Germany, 12.4% of France. The overall Italian production "in SS indexed journals" turns out to be still low and inferior in absolute values compared to other European countries results. This is also confirmed by the fact that the "quota of Italian SS publications compared to the total of the database ISI-Thompson is 3%, against a value EU27 of 5%" (, p. 21). Even recent researches of Cinzia Daraio, dedicated to Italian scientific production as a whole and calculated over a longer period of time (1980-2007, sources WoS and EUROSTAT), confirm that the annual growth rate has a tendency to be higher compared to the main European countries (except Spain). However after 2007, "when the global percentage of Italian journals was 3.5%, a slight decrease in Italian scientific production was noted, with 3.4% in 2008 and 3.3% in 2009" (Daraio, p. 38; Daraio and Moed p. 1380-1392). According to Daraio, "in the year 2000 Italy reached the world average in terms of quality of its scientific production, measured by normalized relative

¹⁷For more information about Spanish situation see González-Alcaide ("Anàlisi del proceso de internacionalización de la investigación española en ciencia y tecnología (1980-2007)").

impact of citations; however its level is lower than the one of the main European countries (Switzerland, Netherlands, Great Britain, Germany, France)" (Daraio, p. 39). In view of the foregoing, it is evident how bibliometric instruments, particularly those taken in consideration on this occasion, do not represent Humanities and Social Sciences disciplinary areas in an adequate way. Alberto Baccini, summarizing the results of Henk F. Moed's research put together in a study dedicated to the analysis of the adequacy of databases for scientific evaluation (Moed), showed that the level of coverage of ISI for Humanities and Social Sciences should be considered "moderate", that is the lowest among all the disciplines (Baccini, p. 88-97, in part. 97). The Italian Agency of Evaluation (ANVUR) takes these limits into account, and in a recent document it limit the recourse to bibliometric analysis to few disciplinary sectors of Humanities and Social area (ANVUR; ANVUR, GEV 11). The validity of criticism towards the application of bibliometric methods to Humanities and Social disciplines cannot justify the avoidance, by many scientific sectors, of evaluation activities. The last observation highlights another aspect: the inadequacy of cultural approach of the majority of professors when it comes to evaluation issues. Baccini points the finger at the poor attention by the Italian university environment when it comes to these problems, and at the poor interest by editorial committees in seeing their own journals included in international bibliographic databases (Baccini, p. 92).

What are the suggestions for the future? On a general level we believe it is a priority to increase the commitment to improve bibliometric instruments, making them more responsive to the characteristics of the Humanistic and Social area products. In this sense, precious indications can be found on the already cited documents elaborated by the Groups of evaluation experts (GEV) of ANVUR, which describe the criteria used for evaluation of research products

subjected by Italian Universities to VQR 2004-2010 (29 February 2012) and the report of Andrea Bonaccorsi (7 March 2012). From the latter text we get an exhortation to experiment indicators which are not exclusively based on citations analysis, for example: indicators of utilization (*journal usage factor*), indicators of availability in catalogues, web based usage indicators, reviews of monographs. On the bibliographic databases specific level, we rather think it is important to act in two directions. Bibliographic instruments (databases, journals classifications and editorial series, etc.) oriented towards Humanities and Social Sciences production in European context should be identified and developed. Among the endeavours already in progress, the following deserve to be pointed out: the Spanish Clasificación integrada de revistas científicas (CIRC), with its second edition (2011/2012) edited by different institutions (CCHS-CSIC, le università di Navarra, Granada, Carlos III di Madrid, la Fundación Dialnet) and independent from national agencies of evaluation CNEAI and ANECA;¹⁸ and the French index, which is currently being adjusted, created by Agence d'évaluation de la recherche et de l'enseignement supérieur (AERES).¹⁹ In the area of cooperation we point out the European reference index for the humanities (ERIH), whose aim is to increase the visibility of the journals in the area of humanities published in all the languages of European Union. ERIH index is run by European Science Foundation (ESF), an institution established in Strasbourg in 1974 with the aim to improve European cooperation in the area of research.²⁰

In this phase we should considerate counterproductive to foresee the abandonment of instruments such as WoS and Scopus. Even

¹⁸<http://epuc.cchs.csic.es/circ/categorias.html>

¹⁹<http://www.aeres-evaluation.fr/Publications/Methodologie-de-l-evaluation/Listes-de-revues-SHS-sciences-humaines-et-sociales>

²⁰<http://www.esf.org/research-areas/humanities/erih-european-reference-index-for-the-humanities/erih-foreword.html>

with their limitations, they remain an international point of reference for evaluation activities. Actions of reinforcement of the position of Italian language journals are to be pursued, perhaps favouring weaker disciplinary sectors within databases. ANVUR, learning from some international experiences, announced some initiatives going in the right direction. We will examine some of them. The national agency of evaluation decided to organise activities to support “candidacy of a solid group of Italian language journals which satisfy editorial requirements accepted at international level for the purpose of indexation in ISI and Scopus” (Bonaccorsi). During the first phase will be systematically verified the requirements for the access to the databases of Italian journals already indicated in group A (and in some cases in group B) by GEV. ANVUR will take into consideration the experience of the Italian journals, in the humanities and social science area, which have already gained access to ISI and Scopus. For this purpose we point out that this essay, devised also as a cluster of data, can provide a better acquaintance with this reality. The same agency, once the list of the journals candidates will be established, will offer logistic and organisational support to handle the talks with the producers of databases. ANVUR will also commit itself to determine the right modality of monitoring the scientific quality of monographs (Bonaccorsi). A study group will be established, in collaboration with Italian Association of Editors (AIE), in order to elaborate “a grid of indicators aimed at affirming the modality with which the editors should manage the submission and the selection of manuscripts”. Even in this case the proposal is positive, and could be extended at the journals’ managements. The certification of the review (peer review) has now become an essential activity. It can be guaranteed through the traceability of all the stages of the evaluation activity: from a request for public funds to a transparent management of the lists of reviewers, to the care

for the anonymous nature of the review process. There are by now standardized procedures and authoritative guidelines, just think of recent *European Peer review Guide* (2011) of the European Research Foundation,²¹ which can be adopted or suggested as reference criteria.

²¹<http://www.esf.org/activities/mo-fora/peer-review.html>

Works cited

- Abramo, Giovanni, Ciriaco Andrea D'Angelo, and Flavia Di Costa. "National research assessment exercises: a comparison of peer review and bibliometrics rankings". *Scientometrics* 89. (2011). (Cit. on p. 2).
- Anegón, Felix De Moya, et al. "Coverage analysis of Scopus: A journal metric approach". *Scientometrics* 73.1. (2007): 53–78. (Cit. on p. 5).
- ANVUR. *Valutazione della Qualità della Ricerca 2004-2010 (VQR 2004-2010). Documento di accompagnamento dei criteri*. 2012. (Cit. on p. 14).
- ANVUR, GEV 11. *Allegato 1. Valutazione bibliometrica (psicologia e la parte di scienze motorie, afferente all'area pedagogica, ma di cultura biomedica)*. 2012. (Cit. on p. 14).
- Baccini, Alberto. *Valutare la ricerca scientifica. Uso e abuso degli indicatori bibliometrici*. Bologna: Il Mulino, 2010. (Cit. on pp. 2, 14).
- Barbieri, Edoardo, ed. "La valutazione della ricerca nelle discipline umanistiche". *La ricerca universitaria e la sua valutazione* 73.1. (2011): 29–35. (Cit. on p. 11).
- Bonaccorsi, Andrea. *Potenzialità e limiti della analisi bibliometrica nelle aree umanistiche e sociali. Verso un programma di lavoro*. 2012. (Cit. on pp. 11, 16).
- Daraio, Cinzia. "Stato della ricerca scientifica italiana nel contesto europeo: un'analisi bibliometrica sul periodo 1980-2009". *La ricerca universitaria e la sua valutazione* 73.1. (2011): 37–47. (Cit. on pp. 13, 14).
- Daraio, Cinzia and Henk Moed. "Is Italian science declining?" *Research Policy* 40.10. (2011): 1380–1392. (Cit. on p. 13).
- De Bellis, Nicola. *Bibliometrics and citation analysis: from the Science Citation Index to cybermetrics*. Lanham: Scarecrow, 2009. (Cit. on p. 2).
- De Robbio, Antonella. *Metodi bibliometrici per la valutazione della ricerca*. 2010. (Cit. on p. 2).
- Deis, Louise and David Goodman. "Web of Science (2004 Version) and Scopus". *The Charleston Advisor*. (2005): 5.21. <<http://www.charlestonco.com/comp.cfm?id=43>>. (Cit. on p. 5).
- Galimberti, Paola. "Verso un nuovo scenario per la valutazione della ricerca". *JLIS.it* 1.1. DOI: [10.4403/jlis.it-16](https://doi.org/10.4403/jlis.it-16). (2010): 87–110.
- Gonzalez-Pereira, Borja, Vicente P. Guerrero-Bote, and Félix Moya-Anegón. "A new approach to the metric of journals' scientific prestige: The SJR indicator". *Journal of Informetrics* 4.3. (2011): 379–91. (Cit. on p. 4).
- González Alcaide, Gregorio, Juan Carlos Valderrama Zurián, and Rafael Alexandre Benavent. "Análisis del proceso de internacionalización de la investigación española en ciencia y tecnología (1980-2007)". *Revista española de Documentación Científica* 35.1. (2012). <<http://redc.revistas.csic.es/index.php/redc/article/view/725>>. (Cit. on p. 13).

- Graziosi, Andrea. "Quanto vale una citazione nella letteratura scientifica?" *Il Sole 24ore*. (2011). <<http://www.ilsole24ore.com/art/cultura/2011-10-14/quanto-vale-citazione-letteratura-181636\PRN.shtml>>. (Cit. on p. 2).
- Jacsò, Peter. "As We May Search—Comparison of Major Features of the Web of Science, Scopus, and Google Scholar Citation-Based and Citation-Enhanced Databases". *Current Science* 89.9. (2005). <http://cs-test.ias.ac.in/cs/Downloads/article_39430.pdf>.
- . "ISI Web of Science, Scopus, and SPORTDiscus". *Online2* 28.6. (2004). <<http://www.jacso.info/pdfs/jacso-isiwos-scopus-sportd-28-6.pdf>>. (Cit. on p. 5).
- La Guardia, Cheryl. "E-Views and Reviews: Scopus vs Web of Science". *Library Journal* 1.15. (2005). <<http://www.libraryjournal.com/article/CA491154.html>>. (Cit. on p. 5).
- Moed, Henk F. *Citation analysis in research evaluation*. Dordrecht: Springer, 2005. (Cit. on p. 14).
- Reale, Emanuela. "La valutazione della ricerca nelle discipline umane e sociali: significato, pratiche, strumenti". *La ricerca universitaria e la sua valutazione*. (2011): 19–28. (Cit. on p. 13).
- Tarantino, Ezio. "Web of science, Scopus, Google scholar: tre database a confronto (un caso di studio)". *Bollettino AIB* 1.2. (2006): 23–32. <<http://www.aib.it/aib/boll/2006/0601023.htm>>. (Cit. on p. 5).

ANDREA CAPACCIONI, Università degli Studi di Perugia.

andrea.capaccioni@unipg.it

GIOVANNA SPINA, Laureata in Storia e scienze della documentazione (LM) Università di Perugia.

spinaj@hotmail.it

Capaccioni, A., G. Spina. "Italian SSH journals in Journal Citation Reports (JCR) and in SCImago Journal Rank (SJR): data and first analysis". *JLIS.it* Vol.3, n.1 (Giugno/June 2012): 4787-1-4787-20. DOI: [10.4403/jlis.it-4787](https://doi.org/10.4403/jlis.it-4787). Web.

ABSTRACT: Aim of this paper is to give a contribution to the analysis of the visibility gave to the Italian Humanities journals from leading bibliographic databases used for national research assessment. In particular, we will focus on the bibliographic databases Web of Science (Thomson Reuters) and Scopus (Elsevier). We believe it is important to develop a deep understanding of the issues related to assessment. We would like to deal with these issues, not yet adequately discussed by Italian scholars, through an essential statement of the main aspects of the problem. Therefore, we propose an examination of detailed lists of Italian Humanities journals included in the selected databases to help the identification of titles. Then is given a critical analysis of the data.

KEYWORDS: Bibliographic databases; Bibliometrics; Research assessment; Scopus; Web of Science

Submitted: 2012-02-13

Accepted: 2012-03-21

Published: 2012-06-01





Qualità e quantità: stato dell'arte della valutazione della ricerca nelle scienze umane in Italia

Paola Galimberti

La valutazione è un processo che richiede chiarezza nelle scelte legate agli obiettivi e alle motivazioni. Un processo che genera conoscenze, ma anche le utilizza, che promuove apprendimento ma anche introduce la propensione verso comportamenti opportunistici. Il tema della valutazione della ricerca, i suoi obiettivi e gli scopi, ha attirato l'attenzione delle comunità scientifiche disciplinari anche e soprattutto nell'ambito delle scienze umane, dove il discorso su quantità e qualità è sempre stato affrontato con un certo distacco, quasi fosse un tema non di diretta pertinenza. Le riflessioni in questo ambito hanno però subito una improvvisa accelerazione con la nomina dell'Agenzia di Valutazione del Sistema Universitario e della Ricerca (ANVUR) nel 2010 e ancor di più con l'avvio dell'esercizio nazionale di valutazione (Valutazione della Qualità della Ricerca (VQR) 2004-2010),¹ suscitando critiche, proposte e richieste di modifiche, ma anche dando avvio ad una vivace discussione sugli indicatori e sulle metodiche da applicare. Nonostante in molti si siano sentiti in diritto di parlare e scrivere di valutazione della

¹<http://www.anvur.org/?q=schema-dm-vqr-definitivo>



ricerca in questi ultimi tempi,² pochi sono davvero titolati a farlo. La valutazione (della ricerca ma non solo) costituisce un'area di studio di carattere trasversale che associa l'approfondimento delle problematiche, teoriche e pratiche, connesse all'esercizio valutativo (perché si valuta, per chi si valuta, come si valuta, che uso si fa e si può fare della valutazione) ai contenuti e alle caratteristiche dell'oggetto e dei soggetti sottoposti a valutazione.

In pratica per poter correttamente valutare un prodotto della ricerca non basta possedere le competenze scientifiche della corrispondente area disciplinare ma è anche necessario avere dimestichezza con le problematiche, ancora una volta teoriche e operative, per poter esercitare detta competenza. E, ovviamente, viceversa. In Italia manca (o meglio mancava fino a qualche tempo fa) una vera cultura della valutazione, manca una tradizione seria di studi sulla valutazione della ricerca, mancano centri di gestione e analisi dei dati che rappresentino il riferimento per chi si occupa di queste tematiche,³ manca un sistema informativo efficace ed efficiente, mancano dati affidabili e completi, il tutto anche in ragione di una domanda che finora non aveva rappresentato una motivazione così rilevante per attivare un circuito positivo in grado di sostenere processi tendenzialmente deboli e contraddittori.

Con l'avvio dell'esercizio nazionale, pur con tutte le contraddizioni e le critiche derivate, le polemiche ma anche i contributi propositivi, si è assistito a un'inversione di tendenza che ha fatto emergere competenze diffuse, interessi ed esperienze poco noti fino a questo momento e che dovrebbero in qualche modo essere ricondotti a unità per poter essere sfruttati appieno.

²Gli interventi su quotidiani o riviste divulgative, su blog e siti dedicati si sono moltiplicati in questi ultimi tempi, alimentando un dibattito vivace e molto partecipato dai diretti interessati.

³Un esempio potrebbe essere l'IFQ di Berlino o il CWTS di Leiden o il Wissenschaftsrat.

Il primo errore che si ravvisa nei contributi che appaiono quotidianamente nei giornali, nei convegni o nei blog è quello relativo alla capacità di correlare la strumentazione scelta agli obiettivi (espliciti e impliciti) della valutazione, ovvero la non distinzione del livello a cui la valutazione si applica e la diversa natura del committente, da cui derivano poi necessariamente oggetto, indicatori e metodiche utilizzati: come se si potessero applicare gli stessi criteri alla valutazione dei Dipartimenti a livello nazionale o a livello di ateneo, alla valutazione dei settori o dei singoli ricercatori. In realtà deve essere ben chiaro che le diverse dimensioni (Ateneo, Dipartimenti, Aree/gruppi di SSD, SSD, individui) non sono direttamente scalabili o cumulabili o confrontabili, in quanto rispondono a criteri e metodi diversi di analisi, senza dimenticare che i fenomeni sotto osservazione sono dinamici e non sempre la rilevazione viene fatta nel momento ritenuto corretto rispetto al problema da affrontare.⁴

Di conseguenza le classifiche che ne derivano sono sempre relative e tendono alla semplificazione di una realtà complessa.⁵ Nel 2010 un gruppo di esperti della DG Ricerca della Commissione Europea su Assessment of University-Based Research (AUBR) ha pubblicato un interessante rapporto in cui, fra gli altri argomenti, s'introduce il concetto di matrice multidimensionale della valutazione della ricerca, da cui emerge chiaramente la complessità del tema in cui

⁴In riferimento all'attività di un Dipartimento, di un gruppo di ricerca, di un ricercatore. E' possibile infatti che la valutazione prenda in esame un periodo di (apparente) inattività a cui può seguire un periodo molto fecondo, o che non si consideri le trasformazioni nelle mission e nelle composizioni delle strutture piuttosto che i condizionamenti esterni.

⁵Questo punto va tenuto in conto nel momento in cui si decide quale uso fare della valutazione.

Come ben ricordato a Padova da Giovanni Pascuzzi (*La valutazione dei prodotti scientifici nell'area giuridica*) la valutazione incide sull'attribuzione di risorse agli atenei, sulle politiche di reclutamento, sullo stato giuridico dei professori e sulla loro retribuzione, sulle procedure concorsuali e sulla scelta dei progetti da finanziare.

hanno un ruolo fondamentale l'unità scelta, lo scopo, le dimensioni dell'output, gli indicatori bibliometrici e altre tipologie di indicatori (Expert group on AUBR). Come si può vedere nella figura 1 gli indicatori bibliometrici sono una parte all'interno della matrice, a cui però nella discussione nazionale e interna alle comunità scientifiche disciplinari è stato attribuito un peso forse eccessivo,⁶ facendo emergere posizioni radicali sia di fiducia che di sfiducia.⁷

Unit of assessment	Purpose	Output dimensions	Bibliometric indicators	Other indicators
Individual	Allocate resources	Research productivity	Publications	Peer review
Research group	Improve performance	Quality, scholarly impact	Journal citation impact	Patents, licences, spin offs
Department	Increase regional engagement	Innovation and social benefit	Actual citation impact	Invitations for conferences
Institution	Stimulate international collaboration	Sustainability & Scale	International co-authorship	External research income
Research field	Promotion, hiring	Research infrastructure	Citation 'prestige'	PhD completion rates

Figura 1: The multi-dimensional research assessment matrix. This table presents a core part of the matrix, not the entire matrix. It aims to illustrate what the matrix looks like. It should be read column-wise: each column represents a different dimension (Moed e Plume)

È necessario ricordarsi che l'università ha ruoli e funzioni diverse, difficilmente cumulabili ed esprimibili con indici sintetici. Ad

⁶Quasi la valutazione della ricerca potesse essere ridotta agli indicatori bibliometrici.

⁷Si vedano al proposito gli interventi sul sito di ROARS (<http://www.roars.it>) e le repliche di ANVUR (ANVUR. Consiglio direttivo; Bonaccorsi).

esempio un ricercatore molto produttivo e di grande impatto può essere un cattivo docente e viceversa o la sua performance può avvenire a scapito di altre funzioni e/o di altri ricercatori. Nella analisi della produzione scientifica di dipartimenti, ricercatori e gruppi bisogna decidere come considerare le aree/tematiche nuove, come bilanciare le distorsioni determinate da comunità scientifiche piccole, chi premiare e come premiare, come aumentare la performance complessiva e come gestire le informazioni e le conoscenze che si generano, come valorizzare i percorsi migliorativi. Se la qualificazione scientifica avviene a scapito di altre funzioni dell'università bisogna vedere come bilanciare ruoli e funzioni rispetto agli obiettivi da perseguire. In tutti questi casi la matrice che si dovrà comporre sarà diversa e diversi saranno gli indicatori utilizzati.

Qualità e quantità

Mentre l'idea della necessità della valutazione dell'università appare a questo punto largamente condivisa e ormai accettata da tutte le comunità scientifiche, meno condivise sembrano essere le finalità⁸ e ancor meno le modalità attraverso le quali essa deve essere realizzata. L'ANVUR ha dovuto predisporre, in pochissimo tempo, una macchina credibile in grado di associare gli strumenti ai fini dichiarati, dovendo in molti casi svolgere un compito di indirizzo e di giustificazione concettuale delle scelte proposte e, in alcuni casi, di sostituzione rispetto a informazioni, competenze e

⁸"La valutazione della ricerca non deve essere percepita come un meccanismo burocratico o censorio, ma come preziosa opportunità per mettere a fuoco aspetti nevralgici della performance delle strutture di ricerca" [...] (Comitato di Indirizzo per la Valutazione della Ricerca); "Tutte le università dovranno ripartire da zero. E quando la valutazione sarà conclusa, avremo la distinzione tra *researching university* e *teaching university*. Ad alcune si potrà dire: tu fai solo il corso di laurea triennale. E qualche sede dovrà essere chiusa" (Fiori, «Laurea DOC. Intervista a S. Benedetto»).

“tools” non sempre disponibili. In pratica una macchina che, anche in ragione delle scelte fatte sulla quantità e rappresentatività dei prodotti da considerare, permettesse di valutare con metodiche e indicatori differenti a seconda delle aree disciplinari una enorme quantità e varietà di lavori di ricerca, con numeri che avrebbero potuto mettere in crisi sistemi che hanno una tradizione ben più consolidata quali quello australiano o britannico.⁹ Nonostante gli ostacoli la VQR ha preso avvio, e benché i più scettici sottolineino il suo carattere di ipoteca sulle valutazioni future (anche individuali), gli errori metodologici, inevitabili dato il poco tempo a disposizione, l’opinabilità delle scelte compiute (alcune decisamente sbagliate), questo primo passo ha generato riflessioni e discussioni importanti per una evoluzione delle pratiche valutative nell’ambito della ricerca, ha stimolato il confronto spesso anche con toni accesi e ha portato il discorso sulla valutazione, prima relegato nei nuclei di valutazione e negli osservatori, sulle pagine dei quotidiani.¹⁰ Alcuni, dati alla mano, sostengono che piuttosto che una cattiva valutazione meglio nessuna valutazione, e consigliano all’ANVUR di fermarsi per poter partire con indicatori e metodi più robusti.¹¹ In questa sede ci limiteremo a sottolineare che l’avvio di questo esercizio di valutazione nazionale ha portato ad una consapevolezza diffusa che prima era completamente assente e questo è un fenomeno positivo.

In Italia le comunità scientifiche afferenti alle aree delle Scienze umane e Sociali, corrispondenti alle aree CUN 10-14,¹² hanno cominciato solo negli ultimi anni (alcuni settori solo in epoca recentissima)

⁹La VQR 2004-2010 riguarderà 195240 lavori di ricerca e 68237 ricercatori, numeri impressionanti rispetto agli altri esercizi nazionali.

¹⁰Si veda ad esempio il post riassuntivo su ROARS (30 aprile 2012), Il dibattito sui rankings di riviste <http://www.roars.it/online/?p=7836>.

¹¹Si vedano ad esempio i post di Alberto Baccini o di Giuseppe de Nicolao su <http://www.roars.it> in merito alla bibliometria dell’ANVUR.

¹²Vedremo con una serie di distinguo almeno per l’area 13.

ad occuparsi di indicatori quantitativi e qualitativi, ma a fronte della scarsa esperienza hanno potuto contare su un'ampia letteratura sul tema (anche in senso critico) e sugli esiti degli esercizi portati avanti negli altri paesi a livello europeo e internazionale (Wissenschaftsrat; KNAW, *Standard evaluation Protocol 2009-2015*; *Quality indicators for research in the Humanities*; Martin et al.) e hanno avuto quindi (per chi ha saputo coglierla) la possibilità di evitare di prendere strade già percorse e interrotte da chi le ha precedute.

Per tradizione e anche per una serie di motivi contingenti – cronica mancanza di dati (Hicks e Wang; Martin et al.), molteplicità dei canali di comunicazione dei risultati della ricerca difficilmente tracciabili dai database citazionali, differenti modalità e tempi citazionali (Huang e Chang; Hellqvist), uso della lingua nazionale – le scienze umane hanno sempre considerato la *peer review*, peraltro spesso esercitata in forme non strutturate, come l'unico strumento di valutazione possibile per certi ambiti disciplinari. Certamente la *peer review* è stata applicata a vari livelli alla valutazione di progetti, persone, gruppi e lavori di ricerca, ma non sempre i criteri utilizzati (dalla scelta dei revisori, alla modalità di revisione, alla modalità di decisione nel caso di pareri discordi) sono stati resi noti in maniera trasparente. Ciò è vero in particolare per le riviste nazionali (il target per la maggior parte degli studiosi delle scienze umane e sociali).

La brusca accelerazione dei processi legati alla valutazione determinatasi con l'operatività dell'ANVUR ha provocato anche l'accelerazione delle procedure di accreditamento delle sedi editoriali (collane, riviste, case editrici), anche se è difficile capire per il momento se quanto dichiarato dalle varie redazioni corrisponda a una reale prassi editoriale. Si è dunque avviata una vivace discussione, non priva di toni polemicici, sulle modalità di conduzione della *peer review* presso le sedi editoriali nazionali, sul confronto con gli standard europei, sulla opportunità di cercare di pubblicare all'este-

ro (o di pubblicare in Italia ma utilizzando la lingua inglese) e sul senso dell'utilizzo e affiancamento di indicatori quantitativi per la valutazione nelle aree 10-14.

Dovendo attuare valutazioni su ampia scala, che coinvolgono strutture e gruppi di ricerca la *peer review* presenta dei limiti, in particolare se non inquadrata attraverso meccanismi trasparenti e riproducibili, relativi alla qualificazione degli esperti¹³ e a database ricchi di informazioni atte a favorire la corretta allocazione tra oggetto d'analisi e *expertise* del revisore. Tali limiti riguardano soprattutto la credibilità del processo, e la possibilità che riesca a fornire risultati utili anche in termini di comparazione e *benchmarking*,¹⁴ e il meccanismo di gestione, oneroso in termini di tempi e costi. Sempre di più quindi si è affermata (questo a dire il vero anche per le aree 01-09) l'idea dell'impiego della *informed peer review*,¹⁵ cioè di una revisione dei pari in cui al valutatore sono resi noti alcuni elementi in grado di supportarne il giudizio o, eventualmente, di correggerlo e in cui i difetti della valutazione qualitativa si equilibrano con quelli derivanti dalla valutazione quantitativa.

Per le scienze dure la *peer review* viene normalmente affiancata, o preceduta, dagli indicatori bibliometrici che però nelle scienze umane non risultano disponibili per una serie di motivi noti, sia legati ai prodotti (e alla loro modalità di produzione) e alle tipologie di lavori di ricerca (saggi, monografie vs. articoli), sia connessi alle modalità

¹³Si veda ad es. la risposta di P.P. Giglioli a I. Colozzi, Presidente del GEV 14, <http://www.roars.it/online/?p=8285#more-8285>

¹⁴Si pensi alla difficoltà di ricondurre giudizi analitici a giudizi quantitativi o le difficoltà derivanti dalle dimensioni ridotte delle comunità scientifiche in alcuni settori che annullano i meccanismi blind caratteristici della *peer review*.

¹⁵Il documento di accompagnamento dei criteri dei GEV (29 febbraio 2012), individua nella *informed peer review* la tecnica scelta per la valutazione da tutti i GEV, ciascuno secondo le proprie peculiarità e a seconda delle tipologie di lavori valutati, http://www.anvur.org/sites/anvur-miur/files/gev/documento_accompagnamento_criteri.pdf

di citazione che ne rendono difficile la raccolta (citazioni a pie' di pagina piuttosto che in fondo all'articolo, citazioni a monografie e saggi piuttosto che ad articoli), o ai tempi di citazione che sono molto più lunghi che per le scienze dure. La struttura stessa della comunicazione scientifica nelle scienze umane, e il minor interesse economico di un mercato esterno alla "ricerca per la ricerca", ha impedito la creazione di database bibliometrici in grado di fornire indicatori quantitativi e di impatto delle pubblicazioni scientifiche, e i lavori e gli studi orientati in tal senso sono rimasti per il momento a livello puramente teorico.¹⁶ Negli ultimi anni c'è stato un proliferare di indici bibliometrici di vario tipo. Legati all'autore, al lavoro di ricerca, alla rivista. Essi appaiono meno costosi della *peer review*,¹⁷ di più facile implementazione e permettono la comparazione fra strutture gruppi e persone. Gli indici bibliometrici permettono però solo di cogliere alcuni aspetti della qualità di una pubblicazione (o di un autore); sono deboli rispetto alla qualità intrinseca. La principale caratteristica che un indice bibliometrico deve avere è la coerenza

- rispetto a banche dati diverse;
- rispetto alle pratiche della comunità scientifica di riferimento.

Moed («The multi-dimensional research assessment matrix») ha sottolineato come con l'avanzare degli studi sugli indici e l'applicazione degli stessi si sia ormai arrivati alla terza generazione di indicatori bibliometrici (tabella 1 nella pagina successiva).

L'elemento determinante ai fini di disporre di indici coerenti è la disponibilità di dati bibliografici accurati. Questi dati, per le

¹⁶Si veda lo studio commissionato da alcune società scientifiche europee (Martin et al.): in un recente convegno a Venezia lo stesso Ben Martin ha dichiarato che non è ancora ben chiaro se i costi per la costituzione di un database bibliometrico per le scienze umane e sociali sarebbero giustificati dai benefici.

¹⁷Anche se il lavoro di pulizia dei dati prima che possano essere utilizzati è piuttosto lungo e oneroso.

Tipo (generazione)	Descrizione	Esempi
Prima generazione	Indicatori di base, facili da ottenere da fonti attestate, ormai disponibili da decenni	Numero di pubblicazioni, numero di citazioni, metriche dell'impatto della rivista
Seconda generazione	Indicatori relativi o normalizzati per area disciplinare	Media di citazioni relativa o normalizzata per l'area disciplinare
Terza generazione	Indicatori basati sulla analisi avanzata del network, utilizzando parametri come la centralità del network	Scimago journal rank, indicatori del "prestigio" (pesa di più una citazione che viene da una rivista di maggior prestigio)

Tabella 1

scienze umane, non sono attualmente disponibili.¹⁸ Anche gli indici bibliometrici hanno attirato una serie di critiche e perplessità. Ad esempio il fatto che sono facilmente manipolabili, il timore che nel lungo periodo l'uso degli indicatori bibliometrici possa condurre a comportamenti adattivi quali la ricerca di pubblicare su riviste considerate eccellenti e per le quali è necessario scrivere di tematiche *mainstream* trascurando linee di ricerca più originali ma più "rischiose" dal punto di vista dell'impatto immediato, oppure il fatto che ci si indirizzi verso tipologie di pubblicazioni considerate più "redditizie" dal punto di vista del riconoscimento (articolo vs. monografia).

In alcuni paesi, ad esempio la Gran Bretagna, l'impiego degli indicatori bibliometrici è stato rifiutato come unica fonte informativa per la valutazione¹⁹

¹⁸Attualmente esistono gli strumenti di raccolta dei dati (le anagrafi locali), ma manca totalmente a livello centrale una politica di gestione e raccolta dei dati. Una policy comune rispetto ai dati (chi, cosa, come, su che arco temporale e secondo quali criteri di validazione) potrebbe modificare radicalmente la situazione (Capaccioni e Spina).

¹⁹Non così in altri paesi: si veda infatti l'International audit of research development and Innovation commissionato dalla repubblica Ceca a Technopolis dove ricercatori del CWTS hanno effettuato una analisi sulla qualità della ricerca basandosi esclusivamente su indicatori bibliometrici <http://www.technopolis-group.com/resources/>

Bibliometrics are not sufficiently robust at this time to be used formulaically or to replace expert verdict in the REF (HEFCE).

Il numero sembra oggettivo, ma non dice nulla sul ruolo dell'autore, l'età, la tipologia del lavoro.

In assenza di indicatori bibliometrici e di basi di dati affidabili ed esaustive, ma con la stessa necessità delle scienze dure di poter condurre una *informed peer review*, nell'ambito delle scienze umane alla peer review degli articoli si è pensato di affiancare altri tipi di indicatori che riprendono il principio secondo il quale il valore del contenuto corrisponde tendenzialmente al valore del contenitore.²⁰ Sono nate dunque da qualche anno, e a cura di soggetti diversi, liste di riviste organizzate per classi di merito. La scelta di partire dalle riviste è da un lato strana, dall'altro è stata una scelta obbligata.

E' strano che si sia partiti dalle riviste perché gli articoli su rivista (con percentuale variabile nelle aree 10-14) non erano e non sono assolutamente il canale privilegiato per la comunicazione dei risultati scientifici. Gli umanisti scrivono e ritengono più qualificanti soprattutto saggi o monografie. È stata dall'altro lato una scelta obbligata perché forse più semplice rispetto a un modello per la valutazione delle collane o delle monografie, che attualmente nel nostro sistema

downloads/reports/1315_CzAudit_FinalRep_3-Quality_of_Research_EM.pdf.

²⁰Questo assunto, valido per chi utilizza indicatori riferiti alle riviste (IF SJR SNIP ecc.), risulta smentito dalla distribuzione delle citazioni all'interno dei diversi journal che assai spesso è asimmetrica (pochi articoli che ricevono il maggior numero delle citazioni, molti articoli che non ricevono alcuna citazione) secondo il principio di Pareto. Tuttavia Bonaccorsi («Potenzialità e limiti dell'analisi bibliometrica nelle aree umanistiche e sociali. Verso un programma di lavoro») sostiene che l'assegnazione di riviste a classi di qualità fornisce una informazione a priori circa il valore atteso della qualità dei singoli articoli. Nelle riviste con alto indice di impatto è più difficile pubblicare, c'è un maggiore rejection rate e la selezione è più severa, ci si aspetta dunque una qualità più elevata degli articoli pubblicati.

editoriale non sono univocamente classificabili.²¹ Già da parecchi anni, a livello internazionale, europeo e di singoli paesi, circolano elenchi che classificano le riviste in fasce di merito (tre o quattro a seconda delle aree scientifiche) sulla base di criteri predeterminati. Queste liste sono nate per vari scopi. Fra i più noti e diffusi, ERIH (European Reference Index for the Humanities),²² ora arrivato alla sua seconda versione, era nato inizialmente con lo scopo di documentare la produzione scientifica europea di qualità lasciando ampio spazio alle pubblicazioni in lingue diverse dall'inglese.²³

Il primo elenco di riviste, molto criticato, è stato pesantemente emendato. Siamo alla fase di revisione quasi completata. I criteri di distribuzione delle classi di merito (che prima utilizzavano le lettere A B e C con l'avvertenza che non si trattava di fasce qualitative) distinguono fra riviste nazionali e internazionali che pubblicano ricerche rilevanti, citate e prestigiose a livello nazionale, europeo o internazionale. La *revised list* pubblicata nel 2011 comprende in questo momento solo 10 delle iniziali 15 categorie.

Aeres (l'Agenzia per la valutazione della ricerca francese) ha creato una lista di riviste²⁴ per le Scienze sociali e umane per fornire ai valutatori

la liste des revues qui, par discipline ou domaine, tiennent une

²¹Nel senso che vi è una grande variabilità di qualità e impostazione non solo all'interno di una casa editrice, ma anche all'interno della stessa collana, per cui accanto a ricerche scientifiche troviamo testi divulgativi e didattici.

²²<http://www.esf.org/research-areas/humanities/erih-european-reference-index-for-the-humanities.html>.

²³"It is also a unique project because, in the context of a world dominated by publication in English, it highlights the vast range of world-class research published by humanities researchers in the European languages". <http://www.esf.org/research-areas/humanities/erih-european-reference-index-for-the-humanities/erih-foreword.html>.

²⁴<http://www.aeres-evaluation.fr/Publications/Methodologie-de-l-evaluation/Listes-de-revues-SHS-sciences-humaines-et-sociales>.

place importante dans la diffusion des résultats de la recherche à l'échelle nationale et surtout internationale.

L' Agenzia francese in una prima fase (2008) aveva anche suddiviso i titoli in fasce di merito; nelle liste aggiornate invece c'è un'unica fascia di merito che comprende tutte le riviste accreditate.

L' Australian Research Council (ARC)²⁵ ha deciso di avviare nel 2008 un censimento delle riviste significative per tutte le aree scientifiche elaborando una prima lista di 19500 periodici divisi in 4 classi (A* A B C) poi sottoposta al giudizio delle comunità scientifiche che hanno fatto integrazioni e cancellazioni. Alcuni articoli dimostrano come nel passaggio dalla lista iniziale a quella delle società scientifiche ci siano stati riaggiustamenti anche molto significativi (Lamp).

Alla fine del lavoro di censimento l'ARC ha deciso che la lista non sarebbe stata utilizzata per ERA 2010 (l'esercizio di valutazione australiano).²⁶ Il ministro per la ricerca e l'innovazione ha accolto le lamentele di molti ricercatori che giudicavano la lista inadeguata e fortemente sbilanciata, sottolineavano i possibili effetti sulle scelte editoriali future che avrebbero soffocato le riviste molto specialistiche e non nel mainstream e soprattutto lamentavano il possibile utilizzo (scorretto) di tale lista per la valutazione dei singoli:

There is clear and consistent evidence that the rankings were being deployed inappropriately within some quarters of the sector, in ways that could produce harmful outcomes, and based on a poor understanding of the actual role of the rankings.²⁷

²⁵<http://www.arc.gov.au>.

²⁶http://www.arc.gov.au/era/era_2012/review_of_era10_ranked_outlet_lists.htm.

²⁷Il ministro aggiunge nella sua comunicazione: "Instead of rankings the Australian system will incorporate journal quality profiles. [...] "the removal

Le liste di riviste in Italia

L'attività di censimento e classificazione in fasce di merito delle riviste ha preso dunque avvio anche in Italia con una orchestrazione meno centralizzata e organizzata che negli esempi sopra citati partendo, come avviene in molti casi, dal basso. Questo ha fatto sì che a fronte di un lavoro serio e accurato (ma autoreferenziale) da parte delle singole società scientifiche che hanno avviato le attività di censimento, i risultati finali siano difficilmente comparabili in quanto i criteri utilizzati, la suddivisione in fasce, l'ampiezza di ciascuna fascia sono state decise in maniera indipendente e difficilmente riconducibile a unità.

Alcune società scientifiche, quelle più orientate verso l'internazionalizzazione, includono nella lista tutte le riviste straniere significative per la disciplina, attribuendo di default alle riviste presenti nei database citazionali (WOS principalmente) la classe di merito più alta (si pensi alla recente lista della Società italiana degli economisti).²⁸ Altre società scientifiche hanno invece classificato e suddiviso in classi di merito solo le riviste italiane partendo dal presupposto che comunque le riviste di lingua inglese, se di valore, sono già incluse nei database citazionali e sono quindi già dotate di indicatori bibliometrici.

Il risultato di queste attività non sono dei ranking in cui le riviste vengono messe in ordine di punteggio, ma appunto delle classi di merito, e quindi rating, di ampiezza variabile a seconda dei criteri definiti a priori dalle società scientifiche stesse e in molti casi validati dalle comunità scientifiche di riferimento. Quando si creano stru-

of the ranks and the provision of the publication profile will ensure they will be used descriptively rather than prescriptively" <http://chronicle.com/article/Journal-Ranking-System-Gets/127737/>.

²⁸<http://www.siecon.org/online/wp-content/uploads/2012/03/Nota-CdP-13.03.2012.pdf>.

menti di questo tipo è necessario conoscerne i limiti, non pensare che la qualità del contenitore sia sempre e comunque trasferibile al contenuto, né che queste liste possano essere strumenti oggettivi, che permettano di superare la soggettività e i limiti dei giudizi dei pari.²⁹ Tutti gli strumenti di misurazione quantitativa, per quanto imperfetti, sono comodi perché consentono una rappresentazione sintetica e semplificata della realtà, ed è questo il motivo del loro successo, soprattutto se il giudizio risultante è condiviso, aggiunge informazione (o non introduce criticità) rispetto a quanto comunemente percepito. In una parola se consente una lettura generale accettando magari di rinunciare ai dettagli e alle anomalie.

Nel corso di questi anni ci sono stati tentativi di formalizzare i criteri di inclusione nelle liste di riviste e di suddivisione in fasce di merito. Un punto importante lo ha segnato il cosiddetto documento Graziosi (6 marzo 2009), che fornisce raccomandazioni a società scientifiche e consulte in relazione ai criteri da utilizzare nella formulazione delle proprie proposte di classificazione di riviste ed editori; in particolare, per le riviste italiane vengono indicate le classi di merito (da A a D) e il peso di ciascuna classe, i criteri di qualità a cui ciascuna rivista valutata deve rispondere. Le raccomandazioni suggeriscono anche possibili modelli di pesi da attribuire alle pubblicazioni presenti nelle riviste nelle diverse classi di merito, nell'ottica di premiare le pubblicazioni nelle fasce superiori e rendere impossibile che l'accumulo di pubblicazioni della fascia più bassa possa portare a risultati paragonabili a quelli di chi ha articoli (magari in numero minore) nella fascia più alta. Per le riviste straniere si prevede invece una suddivisione in tre classi di merito. Il documento

²⁹Distorsioni legate a giudizi soggettivi e opportunistici sono state pesantemente sottolineate nelle liste di discussione ad esempio in merito alla classifica delle riviste di biblioteconomia e bibliografia, che premia solo uno dei due settori, e trascura il fatto che qualsiasi scienza applicata deve misurarsi costantemente con la professione, che è insieme applicazione dei risultati della ricerca e ricerca applicata.

è stato un punto di riferimento importante e ci si sarebbe augurati che anche i Gruppi di Esperti della Valutazione (GEV) della VQR ne riprendessero lo spirito nell'ottica della creazione di liste condivise con le comunità scientifiche. Invece il poco tempo a disposizione, l'estrema frammentazione delle discipline³⁰ e i limiti numerici che i GEV si sono autoimposti hanno fatto dimenticare il tema dei criteri e le scelte fatte, soprattutto in alcuni settori, risultano piuttosto discutibili.³¹

ANVUR ha più volte ripetuto che le liste di riviste verranno utilizzate come supplemento di informazioni per i valutatori, e che saranno più i risultati della VQR a validare le liste che non, viceversa, le liste di riviste a determinare i risultati della valutazione; tuttavia rendere esplicita la modalità del loro utilizzo, sapendo in anticipo come queste liste verranno utilizzate per progetti pilota, al di fuori e indipendentemente dall'esercizio nazionale come avviene in altri paesi, avrebbe forse rassicurato gli animi, e non introdotto turbative in un processo già per sua natura molto complesso.

Senza voler riprendere in questa sede tutti gli argomenti portati in questi mesi pro e contro le liste di riviste, si vuole però sottolineare che lo stimolo verso un adeguamento a standard internazionali (anche nell'ambito di riviste che pubblichino solo nella lingua nazionale) e all'adozione di criteri condivisi di qualità, a partire dalla trasparenza dei meccanismi di revisione, delle liste dei revisori e dei tassi di rifiuto degli articoli e dalle caratteristiche formali delle riviste (metadati, indicizzazione, presenza sul web, presenza di abstract e parole chiave),³² non può che giovare all'editoria scientifica nazionale. Questa particolare fase storica rappresenta un'opportu-

³⁰Cosa che è portato a creazioni di liste quasi per ogni SSD.

³¹Il sito <http://www.roars.it> ha raccolto nel corso di questi mesi la maggior parte delle posizioni sul tema liste di riviste e scienze umane.

³²Si veda a questo proposito il progetto *Latindex* e le caratteristiche editoriali richieste per le riviste, <http://www.latindex.unam.mx/documentos/docu.html>.

nità di sprovincializzazione anche per gli editori italiani e affinché si adottino pratiche ormai comuni in altri paesi. Tutto ciò tenendo conto anche del ruolo divulgativo e culturale rappresentato da molte ricerche di ambito umanistico, e dunque della molteplicità dei canali di comunicazione e dei destinatari.

La valutazione delle monografie

Come già sottolineato, il lavoro di GEV e società scientifiche si è concentrato in prima battuta sul tema delle liste di riviste, anche se (con distribuzione variabile) il canale più diffuso per la comunicazione dei risultati di ricerca restano il saggio o la monografia. Se anche per le scienze sociali e umane, come per quelle dure, in particolare in alcune aree, vale la prassi di far convergere diversi e precedenti contributi in una monografia, per molti settori questa modalità di produzione e divulgazione rimane quella prevalente con le ovvie conseguenze rispetto ai tempi, ai modi e ai condizionamenti esercitati dal sistema editoriale sull'intero processo. Si è aperto in questi ultimi mesi il dibattito anche con gli editori in merito alla creazione di un albo di accreditamento della qualità delle sedi editoriali (Fiori, «L'università dà i voti agli editori»). L'argomento è piuttosto spinoso, per vari motivi: capita spesso infatti che la pubblicazione di saggi e monografie venga fatta a spese degli autori o delle loro istituzioni. Senza voler togliere nulla al rigore degli editori, il fatto che si paghi per pubblicare, e in assenza di politiche di revisione esplicite, fa pensare che le maglie per la validazione siano in questi casi più larghe (o quantomeno meno confrontabili) e che alcuni editori siano disposti a chiudere un occhio sulla qualità di ciò che pubblicano, ovvero che chiudano un occhio sulle pubblicazioni a pagamento per poi riuscire a finanziare pubblicazioni che scelgono di pubblicare rischiando in proprio.

Una classifica delle sedi editoriali pare poco corretta, soprattutto perché su questo piano le logiche del mercato e quelle della repubblica della scienza vanno a scontrarsi. Gli editori cercano di fare buoni prodotti vendibili per rientrare degli investimenti fatti, i ricercatori vogliono invece che i loro lavori siano letti (apprezzati, citati, criticati) dai colleghi. Il tutto in un'area della scienza dove i fondi sono scarsi e le comunità piccole e specializzate, per cui difficilmente si riescono a trovare collane che siano esclusivamente di ricerca. Capiterà più facilmente che una collana pubblichi testi divulgativi accanto a lavori di ricerca di buon livello. Per questo non ha tanto senso dare un voto al contenitore (l'editore o la collana), quanto invece proporre criteri qualitativi minimi da rendersi espliciti. In particolare la composizione del comitato scientifico, le modalità di peer review, un elenco dei revisori. Fra i problemi più urgenti nella valutazione delle monografie c'è la definizione delle caratteristiche formali (ad es. estensione in numero di pagine) e il significato stesso di monografia (per taluni opera di un solo autore, per altri opera che approfondisce un tema rispetto ad un ambito più vasto). C'è la collocazione di alcune tipologie come ad esempio i manuali che spesso hanno anche un impatto scientifico. E c'è il problema di come individuare la diffusione di un'opera.

Un recente filone di studi (Hicks e Wang; Torres-Salinas e Moed) vede l'analisi dei cataloghi delle biblioteche³³ come possibile fonte di dati sulla diffusione di un'opera, equiparando la presenza di un libro in un catalogo di biblioteca ad una citazione. L'analisi, perché possa dare una qualche indicazione, va fatta su cataloghi di biblioteche di ricerca che siano di riferimento per un determinato ambito disciplinare e che abbiano una politica degli acquisti molto precisa e dichiarata.³⁴ Inoltre questo tipo di analisi è abbastanza inefficace

³³*Library Catalog Analysis (LCA)*.

³⁴Il che non è semplice in un periodo in cui i budget delle biblioteche sono in fase

per quanto riguarda i saggi contenuti in un volume miscelaneo o di atti di convegno. Un altro filone di studi affronta invece il tema delle book reviews all'interno delle riviste (Zuccala e van Leeuwen) cercando di sviluppare indicatori utili per la valutazione delle ricerche. Anche qui siamo solo agli inizi e gli strumenti di riferimento sono ancora una volta quelli da cui la maggior parte della nostra ricerca è esclusa (in particolare Web of science), inoltre i tempi attraverso i quali un lavoro stimola recensioni sono piuttosto lunghi e le recensioni possono essere anche delle stroncature, tuttavia forse in questo ambito i dati potrebbero essere più facilmente recuperati. Le recensioni contribuiscono alla definizione della qualità di un lavoro, sono dunque importanti, ma poiché non vengono valutate (almeno non nello schema previsto dall'ANVUR per la VQR, c'è il rischio che nessuno le scriva più.³⁵ Resta poi la possibilità di utilizzare i dati citazionali di Google Books (Kousha e Thelwall) o di Google Scholar. Anche qui, al di là degli entusiasmi favoriti dalla – apparente – “democraticità” del percorso e dalla disponibilità diffusa dell'informazione ivi contenuta, il lavoro di recupero e pulizia dei dati risulta piuttosto oneroso, così come la trasparenza delle fonti; anche qui gli studi sono appena all'inizio, per cui per poter avere risultati minimamente credibili si dovrà attendere ancor parecchio.

Conclusioni

Benché le discipline umanistiche abbiano cominciato solo recentemente (o molto recentemente) ad occuparsi delle tematiche della valutazione della ricerca e delle problematiche connesse, e benché

di contrazione e soprattutto la maggior parte delle risorse è vincolata agli acquisti di periodici e banche dati.

³⁵In alcuni atenei si cominciano a raccogliere le recensioni anche nelle anagrafi della ricerca, e ANVUR stessa ha previsto per la VQR un campo dove segnalare recensioni e menzioni particolari.

manchi una tradizione di studi e di ricerche su questi temi, la nomina dell'ANVUR e l'avvio dell'esercizio nazionale di valutazione ha costretto anche le comunità scientifiche più restie ad avvicinarsi al dibattito sui metodi della valutazione e sugli indicatori, a recuperare velocemente il ritardo rispetto agli altri paesi europei così come ad una approfondita riflessione su come coniugare quantità e qualità nell'analisi della produzione scientifica di aree, gruppi, Dipartimenti e Atenei in una maniera che sia comparabile.

Il primo problema che balza all'occhio è la difficoltà a dotarsi di un supporto informativo condiviso da cui attingere (dati, indicatori etc.), e il fatto che la soluzione a tale problema non possa venire dal processo, peraltro in atto, di progressiva estensione della copertura dei principali database esistenti, nati nell'ambito delle scienze dure, verso questi settori. Temi quali la frammentazione delle aree scientifiche, le diverse modalità di produzione e divulgazione dei risultati, unitamente ai modi e agli strumenti impiegati per il riconoscimento reciproco dei contributi utilizzati dai singoli autori per l'avanzamento culturale e scientifico, per non parlare delle barriere linguistiche e della non sequenzialità temporale delle proposte e delle idee, rendono queste aree sicuramente peculiari in materia di informazioni utili, del loro stoccaggio, del loro trattamento. Da qui lo sforzo per tentare di uscire da una modalità di valutazione esclusivamente centrata sul "giudizio di valore" esercitato da uno o più esperti, sia nelle sedi accademiche, sia in quelle editoriali, sia nella combinazione dei due mondi. Uno sforzo che è ancora in cerca di una credibile proposta.

A rendere ulteriormente problematica tale prospettiva contribuisce la difficoltà a individuare un "mercato" realmente intenzionato a sostenere tale processo e ad affrontare i relativi costi. Quanto è avvenuto in questi mesi nel dibattito sull'esercizio VQR in queste aree ha teso a privilegiare i problemi piuttosto che a evidenziare le soluzioni.

Infatti, se i dati languono, sono ancora più carenti le conoscenze che dovrebbero concorrere a concettualizzare la progettazione delle attività per raccogliere e trattare le informazioni necessarie.

Il secondo punto da sottolineare è, di conseguenza, la necessità dell'avvio di studi e ricerche sulla valutazione nell'ambito delle scienze umane che non siano legati a processi di valutazione in atto, ma che ne siano del tutto svincolati e che permettano di analizzare, anche studiando esperienze di altri paesi, come coniugare quantità e qualità, e che tengano presente il tema dell'impatto (social innovation) che hanno le ricerche in questi ambiti e che richiedono lo sviluppo di indicatori ad hoc.³⁶ Senza peraltro trascurare, in questi studi, l'analisi della concreta possibilità per poter raccogliere le informazioni, per evitare un rischio di astrattezza che costituisce il parallelo alla via di fuga in cui molti cadono ("non sono certo contro la valutazione ma quella che viene proposta non può essere usata per le scienze sociali e umane..."). Una delle possibili ricadute, certamente dipendente dalla credibilità e dalla condivisione degli esiti, che l'esperienza VQR potrà esercitare sulla realtà nazionale sarà costituita dal moltiplicarsi di esperienze locali che si rapportino e confrontino tra di loro, sul modello degli Osservatori sulla ricerca e, magari, una spinta verso un "traino europeo" come già avvenuto più di vent'anni fa per le scienze dure nell'ambito dei lavori, e della sponsorizzazione culturale ed economica, decisi in sede di Commissione e Programma quadro.

Infine una riflessione va avviata sul concetto stesso di qualità della ricerca. Se nel mondo delle scienze dure tale concetto viene associato o al contributo fornito alla soluzione dei problemi, pur nel loro diverso grado di concretezza e di comprensione da parte dei non addetti ai lavori, o alla significatività di innovazione, misu-

³⁶A questo proposito sarà interessante vedere come nel REF si riuscirà a misurare il "social impact" e con quali ripercussioni sulla valutazione delle strutture.

rata sul livello reciprocamente riconosciuto dello "stato dell'arte", peraltro solo parzialmente associabile al superamento di difficoltà concettuali o interpretative, come nel noto caso della fusione nucleare, tra le scienze umane in molti casi il riconoscimento del contributo e della sua importanza avviene con modalità e tempi difficilmente programmabili. Gli avanzamenti incrementali, sicuramente più parametrabili, finiscono per prevalere su quelli radicali, almeno nel breve termine. L'eterodossità e la variabilità nella produttività finiscono col penalizzare il nuovo rispetto al consolidato, o meglio, non favoriscono una generale riconoscibilità della rilevanza del nuovo, tema che costituisce, peraltro, una delle maggiori critiche (l'atteggiamento conservativo) rispetto all'uso della peer review. Serve quindi un rinnovato sforzo di concettualizzazione per poter fare un significativo passo in avanti: e a questo sforzo devono poter concorrere le diverse conoscenze finalizzate a un lavoro comune, come nel caso di chi la ricerca la fa (docenti e ricercatori) e chi ne classifica e ne documenta i risultati (bibliotecari, esperti informatici), affiancati naturalmente da competenze sui diversi aspetti della valutazione. Senza trascurare, anche per questa area della ricerca, le trasformazioni che derivano dai moderni strumenti della comunicazione e dell'informazione, oggi ancora confinati al "supporto" fornito ad un processo produttivo ancora "sequenziale e analogico" ma che appare destinato a "digitalizzarsi" sempre di più, con tutte le conseguenze, difficilmente prevedibili, anche per la valutazione.

Riferimenti bibliografici

- ANVUR. Consiglio direttivo. «La bibliometria della VQR». (2012). <http://www.anvur.org/sites/anvur-miur/files/la_bibliometria_della_vqr.pdf>. (Cit. a p. 4).
- Bonaccorsi, Andrea. «Potenzialità e limiti dell'analisi bibliometrica nelle aree umanistiche e sociali. Verso un programma di lavoro». (2012). <http://www.anvur.org/sites/anvur-miur/files/valutazione_aree_umanistiche_e_sociali.pdf>. (Cit. alle pp. 4, 11).
- Capaccioni, Andrea e Giovanna Spina. «La presenza delle riviste italiane di area umanistica e sociale nel Journal Citation Reports (JCR) e nello SCImago Journal Rank (SJR): dati e prime analisi». *JLIS.it* 3.1. DOI: [10.4403/jlis.it-4787](https://doi.org/10.4403/jlis.it-4787). (2012). (Cit. a p. 10).
- Comitato di Indirizzo per la Valutazione della Ricerca. *Linee guida per la valutazione della ricerca VTR 2001-2003*. 2006. (Cit. a p. 5).
- Expert group on AUBR. *Assessing Europe's University based research, rapp. tecn., European Commission, DG for Research*. 2010. (Cit. a p. 4).
- Fiori, Simonetta. «Laurea DOC. Intervista a S. Benedetto». *La Repubblica*. (2012-02-04): 49. <<http://ricerca.repubblica.it/repubblica/archivio/repubblica/2012/02/04/laurea-doc.html>>. (Cit. a p. 5).
- . «L'università dà i voti agli editori». *La Repubblica*. (2012-05-03): 54. <<http://ricerca.repubblica.it/repubblica/archivio/repubblica/2012/05/03/universita-da-voti-agli-editori.html>>. (Cit. a p. 17).
- HEFCE. «Interim report of the REF bibliometrics pilot exercise». (2009). <<http://www.hefce.ac.uk/media/hefce/content/pubs/2009/rd1309/rd13\09.pdf>>. (Cit. a p. 11).
- Hellqvist, Björn. «Referencing in the humanities and its implications for citation analysis». *Journal of the American Society for Information Science and Technology*. ISSN: 15322882, 15322890. DOI: [10.1002/asi.21256](https://doi.org/10.1002/asi.21256). (2010). <<http://doi.wiley.com/10.1002/asi.21256>>. (Cit. a p. 7).
- Hicks, Diana e Jian Wang. «Towards a Bibliometric Database for the Social Sciences and Humanities». (2009). <http://works.bepress.com/cgi/viewcontent.cgi?article=1017&context=diana_hicks>. (Cit. alle pp. 7, 18).
- Huang, Mu-hsuan e Yu-wei Chang. «Characteristics of research output in social sciences and humanities: From a research evaluation perspective». *Journal of the American Society for Information Science and Technology* 59.11. ISSN: 15322882, 15322890. DOI: [10.1002/asi.20885](https://doi.org/10.1002/asi.20885). (2008): 1819–1828. <<http://doi.wiley.com/10.1002/asi.20885>>. (Cit. a p. 7).
- «Il dibattito sui rankings di riviste». *ROARS*. (2012-04-30). <<http://www.roars.it/online/?p=7836>>.

- KNAW. *Quality indicators for research in the Humanities*. 2011. (Cit. a p. 7).
- . *Standard evaluation Protocol 2009-2015*. 2010. (Cit. a p. 7).
- Kousha, Kayvan e Mike Thelwall. «Google book search: Citation analysis for social science and the humanities». *Journal of the American Society for Information Science and Technology* 60.8. ISSN: 15322882, 15322890. DOI: [10.1002/asi.21085](https://doi.org/10.1002/asi.21085). (Ago. 2009): 1537–1549. <<http://doi.wiley.com/10.1002/asi.21085>>. (Cit. a p. 19).
- Lamp, John W. «Journal ranking and the dreams of academics». *Online Information Review* 33.4. ISSN: 1468-4527. DOI: [10.1108/14684520910985747](https://doi.org/10.1108/14684520910985747). (2009): 827–830. <<http://www.emeraldinsight.com/10.1108/14684520910985747>>. (Cit. a p. 13).
- Martin, Ben, et al. «Towards a bibliometric database for the Social Sciences and Humanities – A European Scoping Project. A report produced for DFG, ESRC, AHRC, NWO, ANR and ESF». (2010). <http://www.dfg.de/download/pdf/foerderung/grundlagen_dfg_foerderung/informationen_fachwissenschaften/geisteswissenschaften/esf_report_final_100309.pdf>. (Cit. alle pp. 7, 9).
- Moed, Henk F. e Andrew Plume. «The multi-dimensional research assessment matrix». *Research trends* 23. (2011). <<http://www.researchtrends.com/issue23-may-2011/the-multi-dimensional-research-assessment-matrix/>>. (Cit. alle pp. 4, 9).
- Moed, Henk F., et al. *Options for a comprehensive database of research outputs in Social Sciences & Humanities. Version 6*. 2009.
- Pascuzzi, Giovanni. *La valutazione dei prodotti scientifici nell'area giuridica*. Intervento presentato alla Giornata di studio su La valutazione della ricerca nelle aree umanistiche, 13 aprile 2012, Padova. (Cit. a p. 3).
- Torres-Salinas, Daniel e Henk F. Moed. «Library Catalog Analysis as a tool in studies of social sciences and humanities: An exploratory study of published book titles in Economics». *Journal of Informetrics* 3.1. ISSN: 17511577. DOI: [10.1016/j.joi.2008.10.002](https://doi.org/10.1016/j.joi.2008.10.002). (Gen. 2009): 9–26. <<http://linkinghub.elsevier.com/retrieve/pii/S1751157708000527>>. (Cit. a p. 18).
- Wissenschaftsrat. «Empfehlungen zur vergleichenden Forschungsbewertung in den Geisteswissenschaften». (2010). <<http://www.wissenschaftsrat.de/download/archiv/10039-10.pdf>>. (Cit. a p. 7).
- Zuccala, Alesia e Thed van Leeuwen. «Book reviews in humanities research evaluations». *Journal of the American Society for Information Science and Technology* 62.10. ISSN: 15322882. DOI: [10.1002/asi.21588](https://doi.org/10.1002/asi.21588). (October 2011): 1979–1991. <<http://doi.wiley.com/10.1002/asi.21588>>. (Cit. a p. 19).

PAOLA GALIMBERTI, Università degli Studi di Milano.
paola.galimberti@unimi.it

Galimberti, P. "Qualità e quantità: stato dell'arte della valutazione della ricerca nelle scienze umane in Italia". *JLIS.it* Vol.3, n.1 (Giugno/June 2012): 5617-1-5617-25. DOI: [10.4403/jlis.it-5617](https://doi.org/10.4403/jlis.it-5617). Web.

ABSTRACT: Il tema della valutazione nelle scienze umane è divenuto di grande attualità soprattutto dopo l'avvio dell'esercizio di valutazione nazionale (VQR 2004-2010). Più ancora che per le scienze dure il dibattito sugli indicatori, sul come (e se) correlare qualità e quantità, sul concetto stesso di qualità della ricerca è molto vivace non solo per le caratteristiche intrinseche della produzione scientifica in queste aree, ma anche per una cronica mancanza di dati bibliografici e bibliometrici. Il contributo fa il punto sullo stato dell'arte, tenendo presente le esperienze, di altri sistemi della ricerca, e tenta di individuare alcune direttrici lungo le quali sarebbe opportuno muoversi in futuro: la creazione di una base di dati bibliografici secondo standard condivisi, la costruzione di un griglia di riferimento per la valutazione delle monografie, un migliore sfruttamento delle nuove tecnologie sia per la distribuzione e validazione dei lavori di ricerca che per la loro valutazione.

KEYWORDS: ANVUR; Bibliometria; Italia; Valutazione della ricerca

ACKNOWLEDGMENT: L'Autrice ringrazia Alberto Silvani per la lettura attenta, per i commenti e i suggerimenti.

Submission: 2012-05-06
Accettazione: 2012-05-21
Pubblicazione: 2012-06-01





Analyzing rare diseases terms in biomedical terminologies

Erika Pasceri

Introduction

A rare disease is a pathological condition with low prevalence and incidence. There are between 6000 and 8000 rare diseases. Many rare diseases are sparsely distributed in some geographic areas and more frequent in others, for reasons linked to genetic factors, environmental conditions that influence the spread of pathogens and the life habits. Thalassaemia, for example, is a relatively common genetic disease in the Mediterranean basin (very common in Southern Italy) and rare in the United States.

A disease or disorder is defined as rare in Europe when it affects less than 5 in 10,000.¹ One rare disease may affect only a handful of patients in the EU, and another touch as many as 245,000. Overall, rare diseases may affect 30 million European Union citizens. In the United States a rare (or orphan) disease is defined as having a prevalence of fewer than 200,000 affected individuals.² Many diseases are much rarer, reaching a rate of one case per 100,000

¹http://ec.europa.eu/health-eu/health_problems/rare_diseases/index_en.htm.

²<http://www.nlm.nih.gov/medlineplus/rarediseases.html>.



persons or more.

Rare disease patients too often face common problems, including the lack of access to correct diagnosis, delay in diagnosis, lack of quality information on the disease, lack of scientific knowledge of the disease, inequities and difficulties in access to treatment and care. These things could be changed by implementing a comprehensive approach to rare diseases, increasing international cooperation in scientific research, by gaining and sharing scientific knowledge about all rare diseases, not only the most “frequent” ones, and by developing tools for extracting and sharing knowledge.

Organizations such as the National Institute of Health (NIH), Office of Rare Diseases Research (ORDR), National Organization for rare Disorders (NORD) and Orphanet provide information to patients and physicians and facilitate the exchange of information among different actors involved in this field by standardization in clinical terminologies, key factors in information retrieval and information exchange.

The ORDR was established in 1993 within the Office of the Director of the NIH, the Federal point of biomedical research. The aim of ORDR is to coordinate and support rare disease research, responding to research opportunities and providing information, promoting international collaboration and interoperation.

Orphanet, on the other hand, was established in 1997 by the French Ministry of Health (Direction Générale de la Santé)³ and the Institut National de la Santé et de la Recherche Médicale (INSERM).⁴ Orphanet maintains a database of information on rare diseases and orphan drugs for all publics and aims to contribute to the improvement of the diagnosis, care and treatment of patients with rare diseases. It includes a Professional Encyclopedia which is a compre-

³<http://www.sante.gouv.fr>.

⁴<http://www.inserm.fr>.

hensive collection of review articles on rare diseases, author-based and peer-reviewed, a Patient Encyclopedia and a Directory of expert Services. This Directory includes information on relevant clinics, clinical laboratories, research activities and patient organizations.

The NORD was founded in 1983 by patients and families who worked together to get the Orphan Drug Act passed. This legislation provides financial incentives to encourage development of new treatments of rare diseases. The purpose of NORD is to supply information about rare diseases, referrals to patient organizations, research grants and all those people that have interest in rare disease. The purpose of NORD is to supply information about rare diseases, referrals to patient organizations, research grants and all those people that have interest in rare disease. It isn't a government agency; it is a non-profit voluntary health agency that exists to serve rare-disease patients and their families. Its primary sources of funding are contributions membership fees.

Objective

The aim of this project is to analyze a specific area of biomedical terminologies, namely rare disease terms. The representation of rare diseases terms has been analyzed in biomedical terminologies such as Medical Subject Headings (MeSH), International Classification of Diseases (ICD)-10, Systematized Nomenclature of Medicine - Clinical Terms (SNOMED-CT) and Online Mendelian Inheritance in Man (OMIM), leveraging the fact that these terminologies are integrated in the Unified Medical Language System (UMLS). It has been analyzed the overlap among sources and the presence of rare diseases terms in target sources included in UMLS, working at the term and concept level.

Material

In this section the resources used in this study are briefly described: the two sources of rare disease terms (ORDR and Orphanet), the four target terminologies (ICD, MeSH, OMIM, and SNOMED-CT) and the UMLS.

The UMLS[®] is a terminology integration system developed at the National Library of Medicine. The UMLS Metathesaurus[®] integrates almost 160 biomedical vocabularies, including the four target vocabularies under investigation (ICD-10, MeSH, OMIM and SNOMED-CT). Synonymous terms from the various source vocabularies are grouped into one concept. Additionally, the Metathesaurus records the relations asserted among terms in the source vocabularies, including hierarchical, associative and mapping relations. Version 2010AB of the UMLS is used in this study. This version contains approximately 2.4 M concepts and 40 M relations.

Source terminologies

The ORDR⁵ publishes a list of rare diseases. This resource does not represent any relations among rare diseases, but groups all the synonyms of a given disorder into a single concept. It maintains a list of 6,857 rare disease concepts (and 11,803 synonyms) on its Web site of which about 800 have extensive information on resources relating to questions by the public. The rare disease concepts are either diseases for which information requests have been made to directly to the Office of Rare Diseases Research, the Genetic and Rare Diseases Information Center (GARD) which is funded by the ORDR and the National Human Genome Research Institute (NHGRI), or NHGRI directly; or (2) diseases from various data sources and those

⁵<http://rarediseases.info.nih.gov>.

that over the last 10 years have been suggested as being rare. The purpose of the Rare Diseases and Related Terms list is to facilitate the distribution of information.

*Orphanet*⁶ provides information about 5,954 rare diseases. Orphanet diseases are organized into a Directed Acyclic Graph. In the Orphanet database, diseases are linked to external reference terminologies, such as ICD10 and OMIM. The Orphanet list of rare diseases comprises 7,715 concepts. We acquired a list of 7,715 preferred terms and 5,224 synonyms. Additionally, Orphanet shared with us the correspondence they established between rare disease concepts and OMIM and ICD10 codes.

Target terminologies

The ICD is the international standard diagnostic classification for all general epidemiological, many health management purposes and clinical use. It is used to classify diseases and other health problems recorded on many types of health and vital records including death certificates and health records. In addition to enabling the storage and retrieval of diagnostic information for clinical, epidemiological and quality purposes, these records also provide the basis for the compilation of national mortality and morbidity statistics by World Health Organization World Health Organization (WHO) Member States. The 10th revision of ICD (ICD-10) is used in this study. It is included in UMLS.

The **MeSH** is a controlled vocabulary developed by the U.S. National Library of Medicine for the indexing and retrieval of the biomedical literature, especially in the MEDLINE bibliographic database. It consists of sets of terms naming some 25,000 descriptors

⁶<http://www.orpha.net>.

in a hierarchical structure that permits searching at various levels of specificity. Version 2011 of MeSH is used in this study. Of note, this version provides partial coverage for the rare disease terms from ORDR. MeSH is one of the terminologies in the UMLS.

The *OMIM* is a comprehensive, authoritative, and timely compendium of human genes and genetic phenotypes developed at John Hopkins University. The full-text, referenced overviews in OMIM contain information on all known Mendelian disorders and over 12,000 genes. OMIM focuses on the relationship between phenotype and genotype. Its terminological component – including clinical synopses – is available through the UMLS.

The *Systematized Nomenclature of Medicine (SNOMED-CT)* is the world's largest clinical terminology developed by the International Health Terminology Standard Development Organization (IHTSDO) for use in electronic health records. It covers most areas of clinical information such as diseases, findings, procedures, microorganisms, pharmaceuticals etc. SNOMED-CT provides a consistent way to index, store, retrieve, and aggregate clinical data across specialties and sites of care. It also helps organizing the content of medical records, reducing the variability in the way data is captured, encoded and used for clinical care of patients and research. The version of SNOMED-CT used in this study is dated July 31, 2010 and is integrated in the UMLS.

In the remainder of this paper, for simplification purpose, ORDR and Orphanet will be named as *sources* and SNOMED-CT, MeSH, OMIM and ICD10 as the *targets*.

Method

UMLS has been used in various data creation, indexing and encoding systems. It accomplishes this by conjoining the sets of

synonyms and concept relationships in its multiple constituent terminologies (Merabti et al.). In this study rare disease terms from the two sources were mapped to the corresponding UMLS concept(s) using an exact match or after normalization. Normalization abstracts away from such unessential differences as case, punctuation, and inflectional variants (e.g., singular vs. plural) and stop words in terms:

Ex. *Glycogen storage disease type 4* → C0017923 (Exact Match);

Ex. *Isolated growth hormone deficiency type IA* → C1849790 (Normalized String).

Because the terms from ORDR and Orphanet are all expected to name (rare) disorders, we restricted the UMLS concepts mapped to disorder concepts through a filter based on the Semantic Group *Disorders* (including such semantic types as *Disease or Syndrome* and *Congenital Abnormality*). This simple filter provides some level of word sense disambiguation.

Results

The first results of the mapping from the sources to UMLS could be summarized in three categories:

1. *Unambiguous concepts*

All the terms of a given concept map to only one Concept Unique Identifiers (CUI):

Ex. **ORD00117** (Acrodysostosis) → **C0220659** (Acrodysostosis);

Ex. **ORPHA001248** (Maxillo-nasal dysplasia) → **C0220692** (MAXILLONASAL DYSPLASIA, BINDER TYPE);

Ex. **NORD00312** (Conn Syndrome) → **C1384514** (Conn Syndrome).

2. *Ambiguous concepts*

The majority of terms of a given concept map to more than one CUIs. There are two more sub-categories:

- *Ambiguous concepts related to granularity issue:*

ORPHA0000	CUI 1 (C0268128)	CUI 2 (C0220987)	CUI 3 (C0268131)
Oroticaciduria	Orotic aciduria		
Orotic aciduria hereditary		Hereditary orotic aciduria	
Orotidylic decarboxylase deficiency			Hereditary orotic aciduria, type 2
Uridine monophosphate synthetase deficiency	—	—	—

Table 1: Example of an ambiguous concept related to granularity issue

As shown in table 1, from a given Orphanet concept, three terms match to three different CUIs and one match to nothing. In this specific case Orphanet grouped together what SNOMED-CT put in hierarchy:

- *Ambiguous concept not related to granularity issue:*

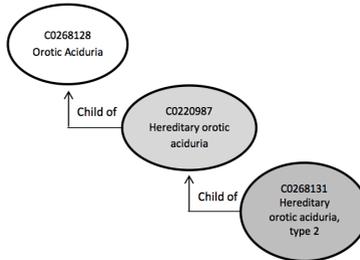


Figure 1

As shown in table 2 on the facing page, from a given Orphanet concept, the terms match to several CUIs, but from UMLS

ORPHA000016	CUI1 (C0339537)	CUI2 (C1844778)
Blue cone monochromatism	Blue cone monochromatism	
Achromatopsia incomplete, X-linked		Achromatopsia, incomplete, x-linked
Achromatopsia, atypical, X linked	—	—
S-cone monochromatism	—	—

Table 2: Example of an ambiguous concept not related to granularity issue.

perspective we don't have any additional information because both terms come from OMIM, so we don't have any information about hierarchical relations.

3. *Unmatched Concepts*

There are some terms from the sources that have no mapping in target sources in UMLS:

- Lateral body wall complex
- Levy-Yeboas Syndrome

The possible explanation for that could be because these are extremely rare diseases (e.g. Lateral body wall complex, approximately 250 cases have been reported in the literature so far) or recently discovered (e.g. Levy-Yeboas Syndrome, discovered in June 2006).

Overall representation in targets

The figure 2 on page 11 shows a part of the overall representation in target sources in the UMLS. On the total number of concepts mapped to UMLS (8,435), we noticed a good representation in the sources we focused the attention:

1.	<i>MeSH</i>	5,663 (67%)
2.	<i>SNOMED-CT</i>	4,192 (50%)
3.	<i>OMIM</i>	3,802 (45%)
4.	<i>ICD10</i>	1,029 (12%)

As shown in figure 2 on the facing page, the blank columns represent those sources that have a very small number of mappings (only one or two). This is because some of them were created for a specific context, e.g.:

- NANDA nursing diagnoses: definitions & classification (NAN)
- Ultrasound Structured Attribute Reporting (ULT)
- Foundational Model of Anatomy Ontology (FMA)

Overlap among sources

Figure 3 on page 12⁷ shows the representation of the overlap among sources. From the ORDR perspective there is 59% of common concepts with Orphanet and 13% with NORD; from Orphanet perspective there is the 43% of common concepts with ORDR and 17% with NORD; and from NORD perspective, there is the 97% of common concepts with ORDR and 92% with Orphanet.

Additional information for a given concept from sources

Among the objectives of this work we set out to find, where provided, additional information for the given concepts from rare dis-

⁷For better details, see <http://leo.cilea.it/index.php/jlis/article/download/SuppFile/4783/5747>.

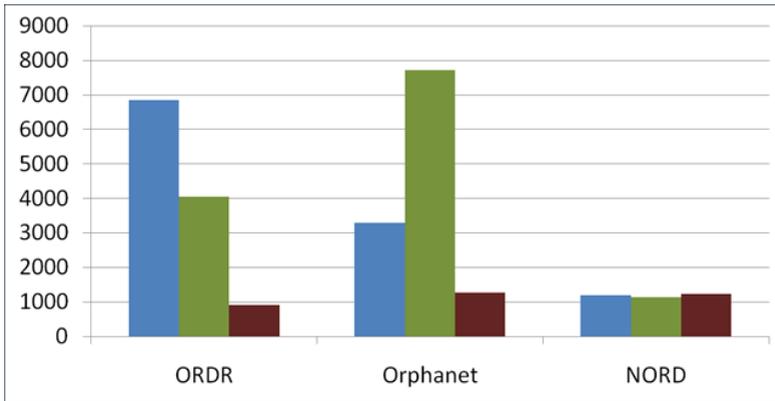


Figure 3: Overlap among sources

eases lists. After analyzing the representation in the target sources, we went deeper in details to find synonyms and more specific terms from target vocabularies. As shown in the example below, from a given concept common in the starting sources, we found that SNOMED-CT can provide additional synonyms and descendants:

Cryptococcosis:

- Torulosis
- Busse-Buschke’s disease
- European blastomycosis
- European Blastomycosis
- Busse-Buschke disease

Additional synonyms provided by SNOMED-CT:

- European cryptococcosis

- Infection by *Cryptococcus neoformans*
- *Torula*

Additional descendants provided by SNOMED-CT:

- Systemic cryptococcosis
- Cryptococcal gastroenteritis
- Cryptococcosis associated with AIDS
- *Cryptococcus* infection of the central nervous system
- Disseminated cryptococcosis
- Hepatic cryptococcosis
- Mucocutaneous cryptococcosis
- Ocular cryptococcosis
- Osseous cryptococcosis
- Pulmonary cryptococcosis

Limitations

In some cases we didn't find any correspondence of terms or concepts in UMLS. This is partly because everything is seen from UMLS perspective; which is because the target sources organize in different ways the terms from their perspectives that makes the difference among the several vocabularies included in UMLS. We also noticed that some concepts not present in UMLS, but probably because there are some diseases that are extremely rare and also because some of them have been recently discovered. If we focus the attention only on Orphanet, maybe we overestimated the percentage of unmapped concepts because in the list of terms there are some that are very general terms as "*rare genetic skin disease*" versus what we have in target sources really specific as "*xeroderma pigmentosus*".

Conclusion

Rare diseases are insufficiently and inconsistently represented in medical terminologies. More than 50% of rare diseases concepts are still not aligned. Automatic approaches can be used to create a draft of the alignment and facilitate the work of domain experts. We found a good representation in target sources in UMLS, especially in the sources where we focused the attention; we also found additional information for the rare diseases concepts. We will share the result with the organizations that work in this particular field so that to enhance the information retrieval. They will provide to review all data with the supervision of clinical experts. This work could be also a feedback to UMLS, for those terms that ORDR, Orphanet and NORD grouped together and UMLS doesn't.

Works cited

- Aymé, Ségolène, et al. "Information on rare diseases: the Orphanet project". *La Revue de médecine interne* 19. (1998).
- Merabti, Tayeb, et al. "Mapping biomedical terminologies using natural language processing tools and UMLS: mapping the Orphanet thesaurus to the MeSH". *Ingénierie et Recherche Biomédicale/BioMedical Engineering and Research* 31.4. (2010): 221–225. (Cit. on p. 7).
- Zhang, Songmao and Olivier Bodenreider. "Alignment of multiple ontologies of anatomy: deriving indirect mappings from direct mappings to a reference". *AMIA Annu Symp Proc.* (2005): 865–868.

ERIKA PASCERI, Università di Udine.
erikapasceri@hotmail.it

Pasceri, E. "Analyzing rare diseases terms in biomedical terminologies". *JLIS.it* Vol. 3, n. 1 (Giugno/June 2012): 4783-1-4783-15. DOI: [10.4403/jlis.it-4783](https://doi.org/10.4403/jlis.it-4783). Web.

ABSTRACT: Rare disease patients too often face common problems, including the lack of access to correct diagnosis, lack of quality information on the disease, lack of scientific knowledge of the disease, inequities and difficulties in access to treatment and care. These things could be changed by implementing a comprehensive approach to rare diseases, increasing international cooperation in scientific research, by gaining and sharing scientific knowledge about and by developing tools for extracting and sharing knowledge. A significant aspect to analyze is the organization of knowledge in the biomedical field for the proper management and recovery of health information. For these purposes, the sources needed have been acquired from the Office of Rare Diseases Research, the National Organization of Rare Disorders and Orphanet, organizations that provide information to patients and physicians and facilitate the exchange of information among different actors involved in this field. The present paper shows the representation of rare diseases terms in biomedical terminologies such as MeSH, ICD-10, SNOMED CT and OMIM, leveraging the fact that these terminologies are integrated in the UMLS. At the first level, it was analyzed the overlap among sources and at a second level, the presence of rare diseases terms in target sources included in UMLS, working at the term and concept level. We found that MeSH has the best representation of rare diseases terms.

KEYWORDS: Rare Diseases; MeSH; Terminology; Text mining; Thesauri

ACKNOWLEDGMENT: This research was supported in part by the Intramural Research Program of the National Institutes of Health, National Library of Medicine (NLM) and in part by the University of Udine. I would like to thank Olivier Bodenreider and Bastien Rance from NLM for technical support and Maurella Della Seta from the Istituto Superiore di Sanità (ISS) that has made this experience possible.

Submitted: 2012-02-10

Accepted: 2012-03-02

Published: 2012-06-01





Le preferenze di selezione verbali: un approccio computazionale

Raffaele Guarasci

Introduzione

L'esperimento effettuato e presentato in questo lavoro mira a fornire una rappresentazione delle preferenze di selezione verbali per la lingua italiana, utilizzando delle metodologie basate su corpora. Per preferenze di selezione si intendono i vincoli imposti dai predicati nella realizzazione dei propri argomenti. L'acquisizione di preferenze di selezione da un corpus, proposta inizialmente da Philip Resnik nel 1993 («Semantic classes and syntactic ambiguity»), si può articolare in due fasi: l'estrazione degli argomenti dai corpora scelti e la generalizzazione delle preferenze di selezione verbali da un'ontologia lessicale.

In questo lavoro vengono utilizzati:

LexIt, lessico di valenza per i verbi della lingua italiana come risorsa lessicale;

MultiWordNet, database lessicale multilingua strutturato in classi semantiche organizzate in modo gerarchico, come ontologia (Magnini e Strapparava).

L'analisi proposta si ricollega a uno degli ultimi modelli sviluppati per generalizzare le preferenze di selezione, l'esperimento effettuato da Sabine Schulte Im Walde («Experiments on the Automatic Induction of German Semantic Verb Classes») sui verbi tedeschi, mirato a ottenere una rappresentazione ad alto livello che generalizzi le preferenze di selezione degli argomenti e fornisca una distribuzione del comportamento dei verbi nella lingua tedesca. Il lavoro effettuato sulla lingua tedesca aveva come obiettivo quello di fornire una generalizzazione del comportamento dei verbi, considerando soltanto le classi più generiche e astratte in cima alla rete semantica. Questo alto livello di generalizzazione limita l'osservazione di alcuni comportamenti specifici di determinati verbi. L'analisi qui proposta mira quindi a superare tale problema e a fornire un livello di rappresentazione più dettagliato, che caratterizzi meglio specificità e preferenze di selezione, considerando nella navigazione della rete semantica di MultiWordNet tutte le classi intermedie, dal momento che le classi generali risultano troppo ampie per caratterizzare in modo differente tutti i verbi. Esprimere le preferenze di selezione nei termini di tutte le classi intermedie della gerarchia permette di far venire alla luce comportamenti più specifici nella scelta degli argomenti, che non potrebbero essere considerati guardando solo le classi generiche.

Le preferenze di selezione

Si può definire "preferenza di selezione" la proprietà di un verbo di preferire o meno argomenti di un particolare tipo semantico. Si può dunque vedere la preferenza di selezione come una sorta di vincolo che opera una restrizione specificando quale siano gli argomenti adatti per un dato predicato. Il concetto di preferenze di selezione o di restrizioni semantiche ha una lunga storia ed è

stato ampiamente trattato sia nella Linguistica generativa (Katz e Fodor; Chomsky), che nella Linguistica computazionale. Uno dei primi approcci per caratterizzare le preferenze di selezione è stato quello di Katz e Fodor, basato sulla più ampia teoria semantica della decomposizione del significato delle parole in features lessicali caratterizzanti. Il classico esempio proposto per la lingua inglese è la parola *bachelor* che può indicare un uomo non sposato, scapolo (features: maschio e umano) o un esemplare di foca maschio privo di compagna (features: maschio e animale). Applicare questo modello ai predicati significa identificare per i predicati delle condizioni necessarie e sufficienti perché siano semanticamente accettabili per essere associate a quell'argomento. Tali condizioni sono rappresentate come funzioni booleane. Questo approccio presenta diversi limiti: dalla difficoltà di identificare univocamente delle condizioni necessarie e sufficienti sempre valide, all'impossibilità di formulare alcuni concetti tramite features binarie.

Modello di Resnik

Il modello elaborato da Philip Resnik nel 1993 è considerato il più quotato modello computazionale per le preferenze di selezione e il punto di riferimento per i lavori successivi. La strategia adottata da Resnik si compone di una rappresentazione tassonomica dei concetti e di una formalizzazione probabilistica delle preferenze di selezione definite nei termini di quella tassonomia, che vengono poi computate e analizzate sulla base di frequenze di co-occorrenza tra i predicati e i loro argomenti. La prima componente del modello deve essere una tassonomia concettuale, una rete semantica, nella quale le classi di concetti devono essere strutturate. La seconda componente del modello deve, invece, fornire una caratterizzazione delle preferenze di selezione nei termini di una relazione probabilistica tra predicati e

classi concettuali, basandosi sull'assunzione che un predicato tende ad associarsi prevalentemente con determinate classi di argomenti.

Implementazione computazionale

Nella sua realizzazione computazionale, il metodo di Resnik utilizza come rappresentazione tassonomica l'ontologia di Wordnet,¹ inserendo nel modello tutti i synset rappresentati in WordNet (mentre altri approcci selezionano solo alcuni topnode, in modo da ottenere una rappresentazione più astratta delle preferenze semantiche). L'algoritmo di Resnik funziona nel seguente modo:

1. si assegna il conteggio di co-occorrenza ai synset di WordNet che contengono un nome come testa lessicale; quando una parola ricorre in più di un synset, la sua frequenza viene divisa per il numero di synset. Nelle varie simulazioni del modello, Resnik utilizza principalmente il *Brown Corpus of American English* (Francis e Kucera), il corpus di riferimento della lingua inglese, e CHILDES (MacWhinney e Snow), un corpus comprendente una serie di interazioni dialogiche con bambini;
2. si estende il conteggio di co-occorrenza a tutti i nodi della gerarchia di WordNet.

Questo metodo è utilizzato per calcolare la prior distribution, $p(\text{classe})$, cioè la probabilità che una classe di WordNet prenda una particolare posizione sintattica, indipendentemente dal predicato, che viene stimata usando soltanto la frequenza della classe costituita a partire dagli argomenti di un dato slot, e la posterior distribution,

¹WordNet è un database semantico contenente 90.000 parole inglesi tra verbi, nomi, aggettivi e avverbi, in cui le parole sono organizzate in classi semantiche. Le parole che condividono un topnode nell'ontologia formano un synset. Un synset comprende un insieme di concetti legati da relazioni di sinonimia, meronimia o altro.

$p(\text{classe} \mid \text{predicato})$, cioè la probabilità di una classe di WordNet nella stessa posizione ma tenendo conto del predicato, stimata tramite la frequenza della coppia predicato-nome. La comparazione tra prior distribution e posterior distribution serve a quantificare quanto una classe semantica si adatti a uno slot predicato-argomento; le classi semantiche che si adattano meglio a un particolare slot hanno le probabilità più alte di co-occorrere con questi. Le preferenze di selezione sono quindi costituite dalla relazione tra prior distribution e la posterior distribution, ovvero tra la probabilità del verificarsi di un argomento a prescindere dal predicato e la probabilità condizionata tra i due. La figura 1 nella pagina seguente, tratta da Resnik («Selection and information: a class-based approach to lexical relationships»), illustra questo approccio: dato un set di classi, C (nell'esempio *Legume*, *Animal*, *Trait*), un predicato, *to grow*, e una posizione sintattica, oggetto diretto, la prior distribution $p(\text{classe})$ è confrontata con la posterior distribution $p(\text{classe} \mid \text{predicato})$. Tralasciando il predicato, $p(\text{class})$, *Animal* tende a ricorrere più frequentemente come oggetto diretto rispetto a *Legume*; tuttavia se si effettuano i calcoli introducendo i predicati, *Legume* diventa molto più frequente di *Animal*.

Modello Schulte Im Walde

Il modello elaborato da Sabine Schulte Im Walde per l'assegnazione di tipi semantici agli argomenti dei verbi della lingua tedesca è basato sul medesimo approccio proposto da Resnik. Per ogni combinazione verbo-frame-slot, la frequenza dei filler nominali è estesa alla gerarchia dei 15 topnode esclusivi di GermaNet (Hamp e Feldweg) (*Creature*, *Thing*, *Property*, *Substance*, *Food*, *Means*, *Situation*, *State*, *Structure*, *Body*, *Time*, *Space*, *Attribute*, *Cognitive Object*, *Cognitive Process*). La frequenze delle parole collegate a più di un concetto viene divisa uniformemente tra questi. La differenza principale

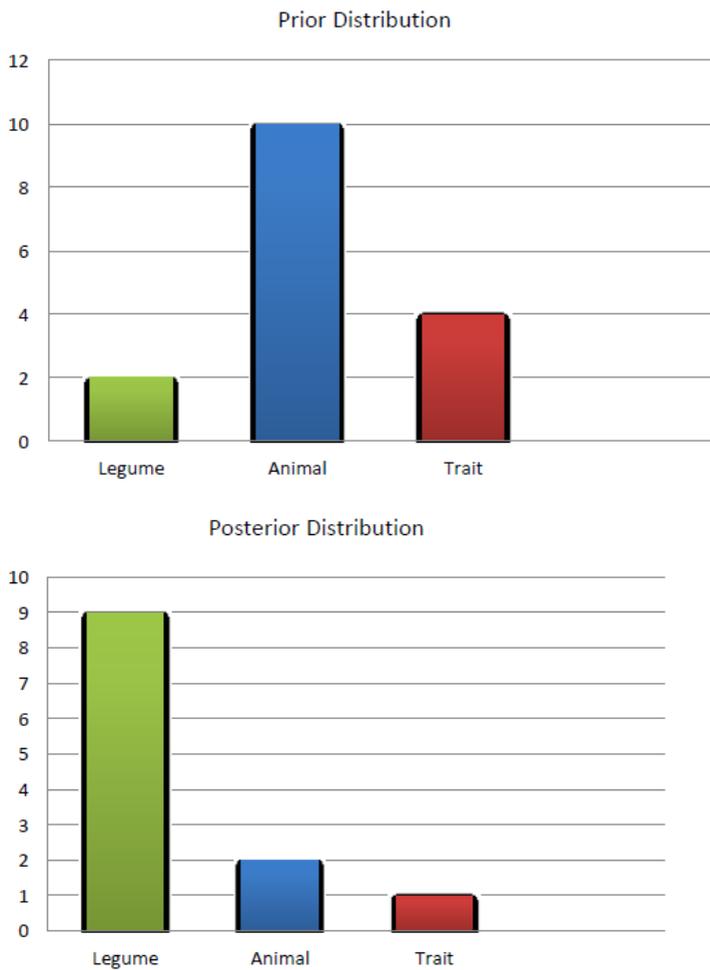


Figura 1

rispetto all' algoritmo di Resnik consiste nel non utilizzare tutti i synset dell' ontologia di WordNet, ma soltanto questi 15 topnode esclusivi. Le frequenze risultanti dei 15 nodi che occorrono in ogni slot sono poi usate per definire una distribuzione probabilistica, normalizzata sulla frequenza totale di co-occorrenza verbo-frame. Oltre a considerare solo questa selezione di synset di GermaNet, l' algoritmo di Schulte Im Walde non apporta sostanziali differenze a Resnik nel calcolo della frequenza di co-occorrenza tra slot e classi, mentre per effettuare il clustering Schulte Im Walde utilizza il valore $p(\text{classe} | \text{predicato})$, diversamente da Resnik che usa una misura proveniente dalla teoria dell' informazione per caratterizzare in modo più ampio le associazioni tra classi semantiche e predicati.

LexIt

LexIt (Lenci, «Carving Verb Classes from Corpora») è una risorsa lessicale per lo studio e l' analisi dei verbi della lingua italiana costruita in modo completamente automatico e basata sul corpus *La Repubblica* e Wikipedia. LexIt rappresenta di fatto la prima risorsa per la lingua italiana che comprenda informazioni distribuzionali sul comportamento dei verbi italiani.² L' obiettivo che questo strumento si pone è quello di descrivere il comportamento dei verbi della lingua italiana sotto il profilo distribuzionale sintattico e semantico, estraendo le informazioni necessarie in modo automatico. Il processo per raggiungere questo obiettivo si articola in diversi sottotask: estrazione dei frame di sottocategorizzazione, assegna-

²La risorsa lessicale più simile a LexIt è rappresentata dal lessico SIMPLE, il quale contiene informazioni sulle preferenze di selezione verbali per le classi semantiche, ma non per filler lessicali o polisemia degli argomenti; inoltre, mentre LexIt è costruito in modo completamente automatico, SIMPLE è stato sviluppato manualmente («SIMPLE: A general framework for the development of multilingual lexicons»).

mento delle preferenze di selezione agli argomenti dei verbi sia come filler lessicali che come classi semantiche, identificazione del ruolo semantico, estrazione automatica delle classi verbali. Il prodotto risultante è una risorsa per lo studio di 3.933 verbi italiani, un database comprendente frame sintattici, filler lessicali e classi semantiche per l'analisi statistica dei verbi italiani. LexIt può anche essere considerato come un lessico di valenza, poiché fornisce, per ciascun verbo, i più significativi pattern sintattici in termini di forza di associazione. La differenza principale tra Lexit e i tradizionali lessici di valenza, quali REDES (Bosque) e il *Wörterbuch der italienischen Verben* (Blumenthal e Rovere),³ consiste nell'essere basato esclusivamente su corpus e costruito in modo completamente automatico. Nella costruzione di Lexit, il corpus scelto è *La Repubblica* (Baroni et al.), sviluppato all'SSLMIT dell'Università di Bologna. Il corpus consta di una collezione di circa 600.000 articoli pubblicati tra il 1985 e il 2000 dal quotidiano *La Repubblica*, per un totale di 386 milioni di parole, divise per tipologia e dominio. La scelta è ricaduta su *La Repubblica* per una serie di motivi: innanzitutto è considerato il corpus di riferimento per la lingua italiana, poi copre un lungo arco di tempo e il linguaggio usato è quello giornalistico, il quale dovrebbe essere rappresentativo dell'italiano standard; inoltre le categorizzazioni di tipo e dominio si prestano ad analisi ulteriori su specifici sottocorpora. Nella fase di preparazione vengono effettuate delle analisi automatiche: il corpus viene prima lemmatizzato e sottoposto a un Pos Tagging tramite ILC-UniPi Tagger (Dell'Orletta), poi a un parser a dipendenze stocastico, DeSR (Bosco, Montemagni e Mazzei). Dal corpus risultante viene ricavato un profilo distribuzionale per ogni verbo trattato e ogni profilo ottenuto viene organizzato in un

³Il *Wörterbuch der italienischen Verben* è l'unico lessico di valenza esistente per i verbi italiani, è basato su un corpus di 50 milioni di parole estratti da "Il Sole 24 Ore" (1989-1990). Il lessico è strutturato in modo che ogni lemma sia articolato in una serie di sensi, ciascuno correlata di esempi.

profilo sintattico e uno semantico. La metodologia usata per costruire LexIt è basata sulle tradizionali misure di associazione applicate allo studio delle collocazioni per valutare la forza di correlazione tra:

- verbi e frame sintattici (estrazione di frame di sottocategorizzazione);
- argomenti verbali e parole che li compongono (identificazione del set lessicale per lo slot di un frame sintattico);
- argomenti verbali e classi semantiche assegnate loro dall'algoritmo.

Profili distribuzionali e semantici per i verbi italiani

Il *distributional profile* di un verbo consiste nell'insieme di informazioni estratte da un corpus per caratterizzare le proprietà distribuzionali del verbo. In letteratura sono stati proposti numerosi metodi per l'acquisizione automatica dei dati, per l'estrazione di frame di sottocategorizzazione, per l'identificazione delle preferenze di selezione verbali (Manning e Schütze; Light e Greiff). Per quanto riguarda l'estrazione di frame di sottocategorizzazione⁴ (SCFs) la caratteristica distintiva di Lexit è il suo approccio automatico e non supervisionato al problema: non si fornisce al modello una lista precostituita di frame sintattici, ma vengono identificati automaticamente le costruzioni sintattiche più frequenti nel corpus. Dopo il processo di estrazione si definiscono gli SCFs del modello. La lista dei suoi frame di sottocategorizzazione (SCFs) di un verbo, ordinati

⁴Per sottocategorizzazione verbale, o valenza verbale si fa riferimento alla capacità dei verbi di scegliere i propri complementi. La struttura di sottocategorizzazione può anche essere definita frame di sottocategorizzazione (SCF), questa fornisce uno strumento per formulare generalizzazioni sul comportamento dei verbi.

secondo la loro rilevanza statistica, definisce il suo profilo. Ogni SCF corrisponde a uno specifico pattern di dipendenze sintattiche di quello specifico verbo ed è formato da un set di slot e identificato da un'etichetta sintetica, ad esempio:

soggetto + complemento introdotto dalla *a* + oggetto diretto = *subj#obj#comp-a*

complemento introdotto dalla preposizione *a* + oggetto diretto = *comp-a#obj*

Nel modello viene anche considerato il pronome riflessivo *si* e il caso in cui un verbo appaia nella forma senza dipendenze, ad esempio nella frase "Il vaso si è rotto" (*subj#si#0*). Il processo di definizione del profilo sintattico avviene selezionando un numero di SCFs tra le combinazioni più frequenti, per ogni verbo si confronta la joint frequency con ogni SCF, basato su pattern estratti automaticamente dal corpus una volta sottoposto al *parser*. Dalla frequenza combinata verbo-SCF si ottiene il punteggio di Local Mutual Information⁵ (LMI) che restituisce una stima della rilevanza statistica di quel SCF per il verbo dato. Nella costruzione di Lexit si parte dal metodo di Resnik per la procedura del conteggio della frequenza e dell'applicazione di una misura di associazione per valutare la correlazione tra un argomento e la classe semantica che occorre insieme ad esso, ma non si utilizzano tutti i synset di WordNet, bensì un gruppo di topnode selezionati che permette di avere una visione generalizzata dei set lessicali. Per l'assegnazione di profili semantici a questi slot sono state implementate una serie di variazioni all'algoritmo di Schulte Im Walde, tra le quali la divisione uniforme tra i sensi assegnati ai nomi nella sezione italiana di MultiWordNet⁶ delle frequenze di

⁵La Local Mutual Information è una variante della Pointwise Mutual Information, che risolve alcuni problemi relativi agli eventi poco frequenti che venivano sottostimati. È una misura comunemente usata nell'analisi delle collocazioni lessicali.

⁶MultiWordNet è un lessico computazionale basato su WordNet. La sezione italiana è allineata con quella inglese: i synset italiani sono creati come corrispondenti di quelli in lingua inglese e le relazioni semantiche sono importate dal database inglese (Pianta, Bentivogli e Girardi).

co-occorrenza di ciascun nome nel ruolo di filler di un dato verbo e il calcolo del valore di associazione effettuato tra ogni combinazione verbo-frame-slot e le top-classi di WordNet. L'approccio usato nella costruzione di LexIt condivide dunque con Schulte Im Walde la selezione di un sottoinsieme di topnode esclusivi usati per un'analisi automatica, mentre le misure di associazione della relazione tra classi semantiche e predicati si basano sul modello precedente di Resnik. I profili semantici assolvono una funzione descrittiva e predittiva: da una parte i set lessicali forniscono una panoramica dei nomi che occorrono nel corpus con un verbo in una determinata posizione sintattica, con una valutazione della loro rilevanza statistica. D'altra parte le preferenze di selezione permettono di generalizzare a partire dalle istanze a delle proprietà astratte degli argomenti dei verbi, consentendo di formulare delle predizioni sugli argomenti non considerati. Un problema centrale in questa fase della costruzione del lessico è stabilire il livello di granularità dell'informazione semantica associata ai filler da MultiWordNet: nella prima fase si selezionano tutte le parole associate ai 24 topnode selezionati, ma questa operazione può creare delle anomalie nei risultati, derivanti dal fatto che MultiWordNet contiene termini anche molto specifici: ad esempio la parola *libro* non rientra solo nella tipologia *Artifact* e *Communication*, ma designa anche una parte dello stomaco dei ruminanti, quindi *Body Part*. Pertanto *libro* ha come associazione principale il verbo *leggere*, mentre *Body Part* è identificata come seconda classe maggiormente associata nello slot di oggetto diretto. Inoltre dalle analisi qualitative viene evidenziato come l'algoritmo non rappresenta correttamente gli usi metaforici, come *leggere la mano* o *leggere le labbra*.

Descrizione dell'esperimento

Il programma realizzato per l'esperimento mira a estrarre le preferenze di selezione dei verbi italiani basandosi sulle informazioni estratte da LexIt e utilizzando la gerarchia di MultiWordNet. LexIt, come detto precedentemente, contiene i profili distribuzionali dei verbi italiani estratti dal corpus "La Repubblica" e *Wikipedia*. Il formato dei dati di output estratti da LexIt utilizzati in questo lavoro è così composto: per ogni verbo sono specificati i frame di sottocategorizzazione, gli slot del frame, i filler nominali e le frequenze per ogni filler. Il formato dei dati estratti da LexIt è dunque composto dalla combinazione verbo-frame-filler-frequenza, come mostrato nell'esempio seguente.

testa verbale + frame	filler nominale	frequenza
avanzare-v%obj%obj	comitiva-s	1
avanzare-v%obj%subj	invasione-s	2
avanzare-v%si#0%subj	candidatura-s	5

Tabella 1: Output LexIt filler

Tali dati estratti da LexIt forniscono precise informazioni lessicali, ma è necessaria una risorsa per generalizzare l'informazione relativa alle preferenze di selezione e associarla alla descrizione dei verbi rispetto a determinati argomenti. WordNet è stato ampiamente utilizzato come fonte per un'informazione a livello abbastanza dettagliato sulle preferenze di selezione (Ribas; Clark e Weir). Pertanto come fonte per l'informazione sulle preferenze di selezione viene utilizzato MultiWordNet che contiene anche la lingua italiana. Come detto sopra, MultiWordNet è organizzato in una gerarchia di synset sinonimi. La figura 2 fornisce una rappresentazione semplificata della gerarchia di MultiWordNet per il nome *salame*.

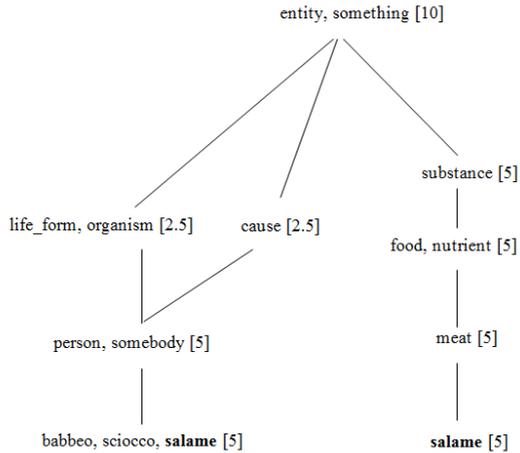


Figura 2: Gerarchia MWN

Come si vede dalla figura 2, il nome “salame” è associato a due sensi: *salame* come sinonimo di persona sciocca e *salame* come tipo di cibo. La gerarchia di MultiWordNet per ogni nome contenuto nei filler dei file estratti da LexIt, viene utilizzata per la costruzione delle preferenze di selezione per quella determinata combinazione verbo-frame-filler a quello specifico livello della gerarchia. L’approccio seguito nella costruzione del programma è simile a quello proposto da Schulte Im Walde («GermaNet Synsets as Selectional Preferences in Semantic Verb Clustering»), e procede in questo modo. Per ogni nome in una combinazione verbo-frame-filler la frequenza è divisa per il numero di sensi del nome e propagata per tutta la gerarchia. Se un synset è collegato a iperonimi multipli, la frequenza viene divisa per il numero di iperonimi, se più synset puntano ad un unico iperonimo, la frequenza viene sommata. Ovviamente la somma

della frequenza dei topnode sarà uguale alla frequenza iniziale della combinazione verbo-frame-filler. Ad esempio se la frequenza del nome *salame* rispetto al verbo *mangiare* in posizione di oggetto diretto è uguale a 10, si assegna a entrambi i synset contenenti il nome un valore pari a 5. I valori di questi nodi si propagano per tutta la gerarchia, dividendosi in caso di nodi multipli e sommandosi in caso più nodi puntino a uno solo. Ripetere l'assegnamento della frequenza per tutti i nomi contenuti negli slot, fornisce una distribuzione della frequenza di tutte le combinazioni verbo-frame-filler estratte da LexIt su tutti i synset di MultWordNet. La novità nell'approccio proposto rispetto all'algoritmo di Schulte Im Walde consiste nel non esprimere le preferenze di selezione soltanto nei termini di una selezione di top-nodes mutualmente esclusivi, ma per tutti i nodi intermedi appartenenti alla gerarchia. Ricapitolando i passi seguiti nell'esecuzione del programma sono i seguenti:

1. per ogni nome in una combinazione verbo-frame-filler la frequenza del nome viene divisa per il numero dei suoi sensi;
2. questa operazione viene propagata a tutta la gerarchia dei synset, se un synset ha più iperonimi, la frequenza viene divisa per il numero di iperonimi, se più synset puntano a un solo iperonimo, la loro frequenza si somma;
3. l'operazione viene ripetuta per tutti i dati estratti da LexIt;
4. viene calcolata la forza di associazione usando la Local Mutual Information per ogni coppia verbo-frame rispetto a un dato synset.

Il formato di output dei dati è coerente con quello utilizzato da LexIt, ogni filler della sequenza verbo-frame-filler contenente un argomento nel file di partenza viene sostituito da tutti i synset che

incontra nel risalire la gerarchia, ad ognuno di questi viene associato il valore di frequenza corrispondente.

testa verbale + frame	synset	frequenza
avanzare-v%obj%obj	accomplishment	7.616
avanzare-v%obj%obj	accusation;accusal	23.375
avanzare-v%obj%obj	activity	1004.9
avanzare-v%si#0%subj	abstraction	70.993
avanzare-v%si#0%subj	act;human_action	218.25
avanzare-v%si#0%subj	affair;occasion	0.3333

Tabella 2: Output LexIt synset

Un elemento importante da considerare è che la lingua di partenza dei synset è l'italiano, in modo da potersi interfacciare con i dati estratti da LexIt, nella navigazione dell'albero si passa alla gerarchia dei synset inglesi, questa scelta è motivata dal voler usare la lingua inglese come metalinguaggio per rappresentare le classi, questo è reso possibile dal fatto che in MultiWordNet i concetti rappresentati sono organizzati in maniera parallela tra le varie lingue, come detto precedentemente.

Analisi dei risultati

Una volta navigata l'intera gerarchia di MultiWordNet i dati risultanti composti da verbo-frame-synset e rispettiva frequenza, vengono valutati utilizzando la Local Mutual Information per stabilire la forza di associazione tra quella coppia verbo-frame e quello specifico synset. Ottenuti i dati ordinati, si procede ad un'analisi qualitativa al fine di mettere in luce fenomeni che non sarebbe stato possibile osservare esprimendo le preferenze di selezione solo in termini dei top-node, come viene fatto attualmente in LexIt. Analizzare

le preferenze di selezione per ogni synset della gerarchia permette infatti di avere una rappresentazione più specifica del comportamento dei verbi nella selezione degli argomenti, basandosi sull'assunzione che le classi ai livelli intermedi della gerarchia possano caratterizzare meglio le specificità del comportamento dei verbi, impossibili da vedere considerando solo le classi in cima alla gerarchia (ovvero i top-nodes), troppo ampie e generiche. Un esempio che dimostra i risultati ottenuti dall'analisi riguarda il verbo *abbagliare* con frame di sottocategorizzazione *#obj* (oggetto diretto), il synset con un più alto maggiore di LMI risulta essere *radiation* nello slot. La tabella seguente e le successive rappresentano i dati nel formato del database di LexIt: verbo % frame di sottocategorizzazione % slot sintattico – synset associato alla fine dell'analisi e forza di associazione.

verbo-frame	synset	LMI
abbagliare-v%obj%subj	radiation	10.586

Tabella 3: Analisi risultati verbo: abbagliare

Se si guardasse soltanto ai top-nodes della gerarchia di MultiWordNet, il verbo andrebbe associato a *natural_phenomenon*, un concetto molto più generico e astratto che può esprimere un insieme di concetti anche molto diversi da *radiation*. È dunque un fenomeno rilevante rispetto all'analisi effettuata solo sui top-nodes e caratterizza meglio il comportamento del verbo per quel determinato frame. Per lo stesso frame si sono riscontrati altri casi degni di nota, alcuni esempi significativi possono essere: il verbo *abbandonare* il cui valore di LMI per il synset *housing;lodging* è molto più alto di quello per *artifact o physical_object* (1.341) e lascia intendere una specificazione maggiore del verbo sugli argomenti preferiti, riferiti a contesti specifici.

Negli esempi seguenti, verbo *avanzare* e *fingerare*, si vede chiara-

verbo-frame	synset	LMI
abbandonare-v%obj%obj	housing;lodging	6.305
abbandonare-v%obj%obj	work	5.696
abbandonare-v%obj%obj	duty	4.091
abbandonare-v%obj%obj	energy	3.768

Tabella 4: Analisi risultati verbo: abbandonare

mente come i synset col più altro valore di LMI cui si associa il verbo risultano molto più caratterizzanti dei corrispettivi top-nodes. In questo caso l'unico top-node cui convergono *proposition*, *inactiveness*, *obedience* e *submissiveness* è *abstraction*, che avrebbe fatto perdere molta informazione sulle preferenze dei due verbi.

verbo-frame	synset	LMI
avanzare-v%obj%obj	proposition	10.421
fingerare-v%obj%obj	inactiveness;inactivity	371
fingerare-v%obj%obj	obedience	221
fingerare-v%obj%obj	submissiveness	217

Tabella 5: Analisi risultati verbi: avanzare, fingerare

Cambiando il frame si osservano altri risultati interessanti, ad esempio restringendo il campo di analisi al frame subj#0 si possono notare altri risultati interessanti, come ad esempio il verbo *avanzare*, il cui secondo synset con maggior valore di LMI, dopo *person*, *individual*, *someone* risulta essere *leader*.

verbo-frame	synset	LMI
avanzare-v%0%subj	leader	253,24

Tabella 6: Analisi risultati verbo: avanzare

Guardando il comportamento del verbo in LexIt, si nota che la classe semantica con cui ha maggior forza di associazione è *Person* (73.7787) e l'iperonimo top-node di *leader* in MWN è *life_form,organism*, quindi il valore alto di LMI risultante da questa analisi restringe il campo e specifica meglio il concetto all'interno di una classe molto più ampia e varia.

Conclusioni e possibili sviluppi

L'analisi effettuata ha permesso di ottenere un livello di generalizzazione più specifico per descrivere il comportamento dei predicati nella selezione dei loro argomenti, effettuare l'analisi e calcolare i risultati per tutte le classi della gerarchia ha consentito di evidenziare delle specificità nelle preferenze di selezione messe in luce dalle classi intermedie di MultiWordNet.

L'analisi, come detto sopra, è effettuata navigando la gerarchia di MWN a partire dai synset italiani estratti da LexIt e passando ai livelli successivi con le classi della gerarchia inglese, perfettamente corrispondente alla gerarchia dei synset italiani. L'esperimento è stato implementato lavorando sui file di testo estratti da LexIt nel formato sopra descritto (verbo-frame-frequenza) e i dati di output per esprimere le preferenze di selezione sono stati mantenuti nello stesso formato (verbo-frame-synset-frequenza), sostituendo al filler nominale il synset della classe corrispondente.

Nonostante questo lavoro possa considerarsi come un esperimento autonomo, indipendente da LexIt, qui utilizzato soltanto come risorsa lessicale-semantica per la distribuzione del comportamento dei verbi italiani, in un'ottica più ampia l'obiettivo e l'evoluzione naturale del programma sono quelli di integrarsi con le funzionalità di LexIt. Un'interrogazione diretta del database di LexIt e la possibilità di arricchirlo con tutte le informazioni fornite dall'analisi delle

preferenze di selezione su tutti i livelli della rete semantica, sono da considerarsi i possibili futuri sviluppi.

Riferimenti bibliografici

- Baroni, Marco, et al. «Introducing the La Repubblica Corpus: A Large, Annotated, TEI(XML)-Compliant Corpus of Newspaper Italian». *Proceedings of Fourth International Conference on Language Resources and Evaluation, LREC 2004*. Lisbon: ELDA, 2004. 1771–1774. (Cit. a p. 8).
- Blumenthal, Peter e Giovanni Rovere. *Wörterbuch der italienischen Verben*. Stuttgart: Ernest Klettverlag, 1998. (Cit. a p. 8).
- Bosco, Cristina, Simonetta Montemagni e Alessandro Mazzei. «Evalita'09 Parsing Task: comparing dependency parsers and treebanks». *Proceedings of the 11th Conference of the Italian Association for Artificial Intelligence Reggio Emilia*. 2009. (Cit. a p. 8).
- Bosque, Ignacio. *Redes: diccionario combinatorio del español contemporáneo*. Madrid: SM Ediciones, 2004. (Cit. a p. 8).
- Chomsky, Noam. *Aspects of the Theory of Syntax*. Cambridge, MA: The MIT press, 1965. (Cit. a p. 3).
- Clark, Stephen e David Weir. «Class-Based Probability Estimation using a Semantic Hierarchy». *Computational Linguistics* 28. (2002). (Cit. a p. 12).
- Cove, Thomas M. e Joy A. Thomas. *Elements of information theory*. New York: Wiley, 1991.
- Dell'Orletta, Felice. «Maximum Entropy for Italian PoS Tagging». *Intelligenza Artificiale* 4.2. (2007): 10–11. (Cit. a p. 8).
- Francis, Nelson e Henry Kucera. *Frequency analysis of English usage*. Boston: Houghton Mifflin, 1982. (Cit. a p. 4).
- Gamallo, Pablo, Alexandre Agustini e Gabriel P. Lopes. «Clustering syntactic positions with similar semantic requirements». *Computational Linguistics* 31.1. (2005): 107–146.
- Grefenstette, Gregory. *Explorations in automatic thesaurus discovery*. Boston: Springer, 1994.
- Hamp, Birgit e Helmut Feldweg. «GermaNet: a Lexical Semantic Net for German». *Proceedings of ACL Workshop Automatic Information Extraction and Building of Lexical Semantic Resources for NLP Applications*. Madrid. 1997. (Cit. a p. 5).
- Katz, Jerrold J. e Jerry A. Fodor. «The structure of a semantic theory». *Language* 39.2. (1963): 170–210. (Cit. a p. 3).

- Kullback, Solomon e Richard A. Leibler. «On information and sufficiency». *Annals of Mathematical Statistics* 22. (1951): 79–86.
- Lenci, Alessandro. «Carving Verb Classes from Corpora». *Word Classes*. A cura di Raffaele Simone e Francesca Masini. Amsterdam-Philadelphia: John Benjamins, 2010. (Cit. a p. 7).
- . «SIMPLE: A general framework for the development of multilingual lexicons». *International Journal of Lexicography* 22.4. (2000): 489–495. (Cit. a p. 7).
- Light, Marc e Warren Greiff. «Statistical Models for the Induction and Use of Selectional Preferences». *Cognitive Science* 26. (2002): 269–281. (Cit. a p. 9).
- MacWhinney, Brian e Catherine Snow. «The child language data exchange system». *Journal of Child Language* 12. (1985). (Cit. a p. 4).
- Magnini, Bernardo e Carlo Strapparava. «Costruzione di una base di conoscenza lessicale per l'italiano basata su WordNet». *Atti del XXVIII Congresso della Società di Linguistica Italiana*. 1994. 415–418. (Cit. a p. 1).
- Manning, Christopher e Hinrich Schütze. *Foundations of Statistical Language Processing*. Cambridge, MA: The MIT press, 1999. (Cit. a p. 9).
- Pianta, Emanuele, Luisa Bentivogli e Christian Girardi. «MultiWordNet: developing an aligned multilingual database». *Proceedings of the First International Conference on Global WordNet*. 2002. (Cit. a p. 10).
- Resnik, Philip. «Selection and information: a class-based approach to lexical relationships». phd. University of Pennsylvania, 1993.
- . «Selectional constraints: an information-theoretic model and its computational realization». *Cognition* 61. (1996): 127–159.
- . «Selectional preference and sense disambiguation». *Proceedings of the ACL SIGLEX Workshop on Tagging Text with Lexical Semantics: Why, What, and How*. 1997. 52–57.
- . «Semantic classes and syntactic ambiguity». *Proceedings of the workshop on Human Language Technology*. 1993. 278–283. (Cit. a p. 1).
- Ribas, Francesc. «On Learning More Appropriate Selectional Restrictions». *Proceedings of the 7th Conference of the European Chapter of the Association for Computational Linguistics*. 1995. (Cit. a p. 12).
- Schulte Im Walde, Sabine. «Experiments on the Automatic Induction of German Semantic Verb Classes». *Computational Linguistics* 32.2. (2006): 159–194. (Cit. a p. 2).
- . «GermaNet Synsets as Selectional Preferences in Semantic Verb Clustering». *Proceedings of the GermaNet Workshop*. 2003. (Cit. a p. 13).

RAFFAELE GUARASCI, Università degli Studi di Pisa.
r.guarasci@studenti.unipi.it

Guarasci, R. "Le preferenze di selezione verbali: un approccio computazionale".
JLIS.it Vol. 3, n. 1 (Giugno/June 2012): 4786-1–4786-21. DOI: [10.4403/jlis.it-4786](https://doi.org/10.4403/jlis.it-4786). Web.

ABSTRACT: Il lavoro mira a fornire una rappresentazione delle preferenze di selezione verbali per la lingua italiana. L'esperimento si ricollega alle metodologie basate su corpora e si articola in due fasi: l'estrazione degli argomenti dai corpora scelti e la generalizzazione delle preferenze di selezione utilizzando un'ontologia lessicale. Le risorse utilizzate sono: LexIt, un lessico di valenza per i verbi italiani, come risorsa lessicale, e MultiWordNet, come ontologia. L'obiettivo è fornire un livello di rappresentazione dettagliato del comportamento verbale navigando l'intera rete semantica e facendo emergere comportamenti più specifici nelle preferenze di selezione degli argomenti verbali.

KEYWORDS: Linguistica computazionale; LexIt; MultiWordNet; Selezione verbale; Verbi

Submission: 2012-02-12
Accettazione: 2012-03-02
Pubblicazione: 2012-06-01





The *International Cataloguing Principles* and their future

Giuliano Genetasio

The *Statement of International Cataloguing Principles (ICP)* was published in February 2009 on the International Federation of Library Association (IFLA) website. The printed version of the ICP was published in August 2009. It includes the original text of the *Statement* and translations in nineteen languages.

From the IME ICC to the ICP

The ICP are the result of eight years work by the IME ICC, IFLA Meeting of Experts on an International Cataloguing Code. The idea for the IME ICC was born in 2001, forty years after the publication of the *Paris Principles*, the statement of principles adopted by the International Conference on Cataloguing Principles held in Paris in 1961. Natalia N. Kasparova, IFLA Cataloguing Section, proposed the convening of an international conference to discuss the *Principles* in the light of the significant changes which had taken place in the cataloguing environment. The idea was taken up by the IFLA Cataloguing Section, the IFLA National Libraries Section and the Deutsche Bibliothek, which decided to organize a series of meetings between 2003 and 2007. The purpose of the IME ICC was to formulate a set of principles to facilitate the international exchange



of bibliographic data and to promote standards for bibliographic records. This was based on three goals:

1. to develop a statement of international cataloguing principles designed for the current cataloging environment;
2. to harmonize national cataloguing codes;
3. to suggest a set of guiding rules for an international cataloguing code (ICC).

The first meeting (IME ICC1) was held in Frankfurt am Main 28 to 30 July 2003. It was attended by 52 cataloguing experts from 26 European countries, plus some representatives from the United States and Australia. The meeting opened with a comparative survey of the European codes aimed at verifying their degree of adherence to the *Paris Principles*. The comparison revealed general adherence to the *Principles*, with a few exceptions.¹ Several background papers on important cataloguing issues were presented at the meeting.² Other noteworthy papers dealt with the International Standard Bibliographic Description (ISBD), Functional Requirements for Bibliographic Records (FRBR) and Virtual International Authority File (VIAF). An entire day was devoted to discussions from the working groups on specific cataloging issues: personal names (Group one), corporate names (Group two), serials (Group three), multipart resources (Group four), uniform titles and general material designation (Group five). The objective was to determine whether and to what extent it was possible to harmonize the different national cataloguing rules and possibly make recommendations

¹E.g., rules for the choice and form of headings for corporate bodies, and headings for multipart resources.

²Access points, general material designation, uniform titles, corporate bodies, main heading, bibliographic units in the digital context, language of bibliographic records, multipart resources, serials and continuing resources.

for the future ICC. During the Frankfurt meeting the first draft of the principles, *Principles for Library Catalogs and Other Bibliographic Files* (PLC), written by Monika Münnich, Hans Popst and Charles Croissant, was also presented. The PLC were designed for the online environment, taking into consideration the distinctive features of OPACs: navigation, browsing, Boolean search, access points, authority records. These issues were addressed from both the cataloguing and the information retrieval perspective. The *Principles for Library Catalogs* were a transition text: they were intended to replace the *Paris Principles*, which they still reflected in terms of structure, and they constituted the first textual basis of the ICP. The IME ICC1 was the most important meeting of the series, both because it laid the foundations for the *Statement of International Cataloguing Principles*, and because some European and U.S. experts (Barbara B. Tillett, chair of the IME ICC Planning Committee, Mauro Guerrini and Elena Escolano Rodríguez) who participated in IME ICC1 also participated in all the subsequent meetings. The revision of the draft principles began immediately after the IME ICC1, leading to the first official draft of the *Statement of International Cataloguing Principles* (ICP2003 or *Frankfurt Principles*) in December 2003. The new title refers more directly to the *Paris Principles*. The transition from catalogue to cataloguing principles indicated a change of identity of the draft principles. Unlike the PLC, the 2003 *Statement* is addressed to the entire information community (archives, museums, etc., in addition to libraries). The 2003 *Statement* no longer refers to cataloguing codes, and deletes the section on OPAC displays. Further meetings were held as follows: the IME ICC2 (Buenos Aires, August 2004, 45 participants) – the *Recommendations from the IME ICC2 for the International Cataloguing Code rules and Glossary*, a document which was intended to form the basis of the future ICC, emerged from this meeting; the IME ICC3 (Cairo, December 2005, 59 participants); the

IME ICC4 (Seoul, August 2006, 56 participants); the IME ICC5 (Pretoria, August 2007, 28 participants). An informal meeting was then held in Quebec City in August 2008. The *Statement of International Cataloguing Principles* assumed its final form in December 2008. It was published online in February 2009 and presented at the 75th IFLA Congress in Milan.

The text of the *Statement*

There is a line of continuity between the *Paris Principles* and the 2009 *Statement of International Cataloguing Principles* – this is testified by some passages which have remained substantially identical. However, there are also significant differences between the two texts. The ICP are designed for electronic and online catalogues. They aim to apply to any type of bibliographic resource, to any kind of library (not just research libraries, as was the case for the *Paris Principles*) and even to other information communities. The *Paris Principles* and their theoretical bases (Charles A. Cutter, Seymour Lubetzky, Eva Verona) remain an important reference for the ICP, but they are complemented by new references: the IFLA's conceptual models Functional Requirements for Bibliographic Records (FRBR) and Functional Requirements for Authority Data (FRAD), the work of Elaine Svenonius and the ISBD. The ICP extend the *Paris Principles*, which only concerned the choice and form of the heading, to every aspect of descriptive cataloging. They include directions for authority records and express their intention to apply to subject indexing. The ICP extend the finding (to find a book of which either the author, title or subject is known) and collocation (to find all books by a particular author, on a particular subject or kind of literature) functions of the *Paris Principles* with the functions of FRBR (Find, Identify, Select and Obtain a bibliographic resource), in a perspective that

emphasizes the centrality of the user. The ICP combine guiding rules for cataloguing codes, already present in the *Paris Principles*, with some directions for OPAC search capabilities. Finally, the ICP update the terminology: they drop the term "heading" in favour of "access point". Wherever possible, they prefer the term "bibliographic data" to "bibliographic records", in line with the trend of current catalogs, that focus on individual data rather than on bibliographic records. Let us look more closely at the text. After the *Introduction* and the *Scope*, the 2009 *Statement* introduces a set of *General Principles* derived from the work of Svenonius. The principles represent a synthesis of Anglo-American cataloguing theory and S.R. Ranganathan's thought, in addition to some general logical principles derived from Gottfried Wilhelm Leibniz. The most important principle is the convenience of the user. The section *Entities, Attributes and Relationships*, based on FRBR and FRAD, summarizes the different aspects of the object of cataloguing. It therefore constitutes an important change from the *Paris Principles*, which only spoke of "book" and "work". The section *Objectives and Functions of the Catalogue* is also derived from Svenonius, who merges FRBR functions with those of the *Paris Principles*, adding the navigation function: "to navigate within a catalogue and beyond (that is, through the logical arrangement of bibliographic and authority data and presentation of clear ways to move about, including presentation of relationships among" entities. As for the *Bibliographic Description*, the ICP state that each manifestation should have a separate description. The bibliographic description should be based on the item (copy) as representative of the manifestation (publication) and should conform to an international standard (ISBD, as explained in a note). *Access Points* is the central and largest section of the ICP. It includes directions for the choice and form of access points, as well as for authority records. The 2009 *Statement* no longer speaks of headings: neither

the term "main heading", conceptually and technically obsolete, nor the term "uniform heading" are used. Uniform headings are replaced by "authorized access points", which can be expressed in the original language or not – uniformity is retained locally but not internationally, in line with the course inaugurated by MLAR, Minimal Level Authority Records, and VIAF. The basis for access points for personal names is no longer the person, but the bibliographic identity (persona) assumed by the author in one or more resources (e.g. "Charles Lutwidge Dodgson" for the mathematical works but "Lewis Carroll" for *Alice*). The *Foundations for Search Capabilities* section provides guidance on what to include in OPAC interfaces and records. The 2009 *Statement* includes a *Glossary*, whose first version was published separately in 2004. The *Glossary* defines all the terms used in a specific way within the ICP. In addition to preferred and non-preferred entries, it includes a section of terms no longer in use, i.e. terms used in earlier versions of the *Statement* but then deleted.

Unsatisfactory aspects of the ICP

Some aspects of the 2009 *Statement* are not entirely satisfactory. The first is the ambiguity of the ICP objectives, which derives directly from the duality of goals of the IME ICC. The ICP are presented as a set of cataloguing principles on the one hand and as a set of directions for the construction of an international cataloguing code on the other. This is connected to the second unsatisfactory aspect of the *Statement* – the apparent absence of an overall theoretical vision. Several completely different levels coexist within the ICP: principles of a general nature (*General Principles*); conceptual models (*Entities, Attributes and Relationships, Objectives and Functions of the Catalogue*); cataloguing rules (*Bibliographic Description, Access Points*); and directions for OPAC interfaces and displays (*Foundations for Search*

Capabilities). Furthermore, a significant part of the ICP is derived from other authors or texts – Svenonius (*General Principles, Objectives and Functions of the Catalogue*), FRBR (*Entities, Attributes and relationships, Objectives and Functions of the Catalogue*), ISBD (*Bibliographic Description*) – without the reworking that this operation requires. The decision to base the IME ICC working groups – whose work has in some cases (seriality, multipart resources, general material designation) been absent from the final version of the *Statement* – on particular cataloguing issues has perhaps contributed to the lack of an overall theoretical vision. Other areas for improvement in the *Statement* are: the lack of directions for the management of local variants; the treatment of bibliographic description, which is only hinted at in the text; the emphasis on the convenience of the user, which is too vague to be useful: if nothing else the identity of the user should have been specified – the local or remote user, the final or professional user (catalogers, reference librarians, etc.). Finally, a number of points are underdeveloped. The *Statement* declares its intention:

1. to apply to all kinds of bibliographic resources;
2. to apply to all aspects of cataloguing, including subject cataloguing;
3. to be suitable for archives, museums, and other communities.

But these statements do not result in anything specific. The first point is pursued only through the adoption of a more generic terminology – “resource” instead of “book”, “creator” instead of “author”, etc. The second point remains a mere aspiration: potentially important reference texts, such as the SHLs issued by the IFLA in 1999, have been neglected. As for the third point, the language and conceptual references of the ICP are still cataloguing-specific, not archival or museum-specific.

The future of the ICP

The IFLA Cataloguing Section has announced its intention to continue the work on the ICP through the revision of the 2009 *Statement* and further development of the ICC. The IFLA Cataloguing Section has produced several documents that provide information about the future of the ICP. The *IME ICC 2008 Resolution* is a declaration included in the final part of the *Statement*. The *Resolution* stated its intention to continue to revise the ICP in relation to the developments in cataloguing and in the IFLA's conceptual models. Other important information on the future of the ICP came from the IFLA Cataloguing Section *Strategic Plan 2009/2011*, whose first goal was to "promote the development and maintenance of the Statement of International Cataloguing Principles". The *Strategic Plan 2009/2011* confirmed the need to revise the ICP on a five-year basis. This intention is also confirmed by the *Strategic plan 2011/2013*, which states its intention to form a working group to assess the expediency of clearly separating the specific cataloguing rules from the principles and of expanding "the statement further towards becoming" a complete cataloguing code. These objectives were confirmed by the IFLA Cataloguing Section *Annual Report 2011* and by the *Minutes* of the Cataloguing Section meeting at the 77th IFLA Congress in Puerto Rico. During the Congress, the working group for the revision of the ICP was formed (members: Agnese Galeffi, Dorothy McGarry, and David Reser). The revision should begin in 2014. How should these statements be evaluated? The proposal to revise the ICP and, in particular, the idea of separating the principles from the cataloguing rules, is sound and desirable, given that some aspects of the Statement are unsatisfactory. The idea of creating the ICC requires more caution. While it is remarkable, this idea runs into several difficulties: the preparation of the RDA, *Resource Description and Access*, the new international cataloguing standard, and, in Italy, of the REICAT,

Regole italiane di catalogazione, the new Italian cataloguing rules, had already started by the time IME ICC's work began. Who therefore were the ICP directions for the construction of cataloguing codes directed at? The authors of the new rules (already underway)? Or a working group with the task of drawing up another international cataloguing code? Despite the RDA mentioning the ICP as one of their sources, which might suggest that the RDA are themselves the desired ICC, the recent documents from the IFLA Cataloguing Section support the second hypothesis (an international cataloguing code drawn up by an IFLA working group). The drafting of an international code is a challenging undertaking, hardly achievable by an IFLA working group without outside help and international consensus. A first attempt to draw up the ICC, started with the IME ICC2 *Recommendations*, has not been followed up. It is currently difficult to obtain an international consensus, since the RDA have the support of some of the most important libraries (the Library of Congress, U.S. National Agricultural Library, National Library of Medicine, British Library, Libraries & Archives Canada, Deutsche Nationalbibliothek and the National Library of Australia will officially adopt the RDA by the first quarter of 2013) and library associations, and will be used in a significant part of Europe. The RDA carry on the tradition of the world's most widespread cataloguing rules (AACR2) and are a conceptually revolutionary cataloguing code, based on the IFLA's conceptual models.³ Only should the RDA fail to become the de facto ICC – at the moment a remote hypothesis – would the idea of drawing up a cataloguing code become realistic. In all other cases, it seems more appropriate to reject the idea of drawing up the ICC, and to aim towards making the ICP a space for theoretical reflection,

³Cfr. Library of Congress Announces Its Long-Range RDA Training Plan (Updated March 2, 2012), http://www.loc.gov/catdir/cpsol/news_rda_implementation_date.html; see also the presentations of the EURIG-JSC Seminar on RDA, <http://www.slaint.org.uk/eurig/documents.htm>.

an ideal term of reference for all cataloguing codes, beginning from the RDA and their future updates.

ICP bibliography

Bianchini, Carlo. "Rassegna critica dei recenti sviluppi della catalogazione in ambito internazionale." *Bollettino AIB* 49.1 (2009): 35-45. Print. Also available on the web: <<http://www.aib.it/aib/boll/2009/0901035.htm>>

Bianchini, Carlo, Pino Buizza and Mauro Guerrini. "Verso nuovi principi di catalogazione. Riflessioni sull'IME ICC di Francoforte." Guerrini, Mauro. *Verso nuovi principi e nuovi codici di catalogazione*. Con Carlo Bianchini, Pino Buizza, Carlo Ghilli, Antonella Novelli, Lucia Sardo. Milano: Sylvestre Bonnard, 2005. 35-65. Print. Also available on the web: <http://eprints.unifi.it/archive/00001800/01/Guerrini_Tillett_IME_ICC1_Francoforte.pdf>

Bianchini, Carlo e Mauro Guerrini. "From Bibliographic Models to Cataloging Rules: Remarks on FRBR, ICP, ISBD, and RDA and the Relationships Between Them." *Cataloging & Classification Quarterly* 47.2 (2009): 105-124. Print. Also available on the web: <<http://eprints.unifi.it/archive/00002177/01/01639370802561674.pdf>>

Creider, Laurence S. «A Comparison of the Paris Principles and the International Cataloguing Principles». *Cataloging & Classification Quarterly* 47.6. (2009): 583–599. Print.

—. «The Development of the Statement of International Cataloguing Principles». *IFLA Cataloguing Principles: The Statement of International Cataloguing Principles (ICP) and its Glossary*. In *20 Languages*.

Edited by Barbara B. Tillett e Ana Lupe Cristán. München: Saur, 2009. 14–24. Print.

Guerrini, Mauro. «In Praise of the Un-Finished: the IFLA Statement of International Cataloguing Principles (2009).». *Cataloging & Classification Quarterly* 47.8. (2009): 722-740. Print.

—. «I principi internazionali di catalogazione (ICP). Un'opportunità o un'occasione mancata?». *Biblioteche oggi* 29.9. (2011): 5–11. Print. Also available on the web: <<http://www.bibliotecheoggi.it/content/n20110900501.pdf>>.

Guerrini, Mauro e Giuliano Genetasio. *I Principi internazionali di catalogazione (ICP). Universo bibliografico e teoria catalografica all'inizio del XXI secolo*. Milano: Editrice Bibliografica, 2012. Print.

International Federation of Library Associations and Institutions. *IFLA Cataloguing Principles: Steps towards an International Cataloguing Code. Report from the 1st IFLA Meeting of Experts on an International Cataloguing Code, Frankfurt, 2003*. Edited by Barbara B. Tillett, Renate Gömpel and Susanne Oehlschläger. München: Saur, 2004. Print.

—. *IFLA Cataloguing Principles: Steps towards an International Cataloguing Code, 2. Report from the 2nd IFLA Meeting of Experts on an International Cataloguing Code, Buenos Aires, Argentina, 2004*. Edited by Barbara B. Tillett and Ana Lupe Cristán. München: Saur, 2005. Print.

—. *IFLA Cataloguing Principles: Steps towards an International Cataloguing Code, 3. Report from the 3rd IFLA Meeting of Experts on an International Cataloguing Code, Cairo, Egypt, 2005*. Edited by Barbara

B. Tillett, Khaled Mohamed Reyad and Ana Lupe Cristán. München: Saur, 2006. Print.

—. *IFLA Cataloguing Principles: Steps towards an International Cataloguing Code, 4. Report from the 4th IFLA Meeting of Experts on an International Cataloguing Code, Seoul, Korea, 2006*. Edited by Barbara B. Tillett, Jaesun Lee and Ana Lupe Cristán. München: Saur, 2007. Print.

—. *IFLA Cataloguing Principles: Steps towards an International Cataloguing Code, 5. Report from the 5th IFLA Meeting of Experts on an International Cataloguing Code, Pretoria, South Africa, 2007*. Edited by Barbara B. Tillett, et al. München: Saur, 2008. Print.

—. «Statement of International Cataloguing Principles». *IFLA Cataloguing Principles: The Statement of International Cataloguing Principles (ICP) and its Glossary. In 20 Languages*. Edited by Barbara B. Tillett and Ana Lupe Cristán. München: Saur, 2009. 27–40. Print. Also available on the web: <http://www.ifla.org/files/cataloguing/icp/icp_2009-en.pdf>.

International Federation of Library Associations and Institutions. Cataloguing Section. «Annual Report 2011». (2012). Web: <http://www.ifla.org/files/cataloguing/reports/annual_2011.pdf>.

—. «Strategic Plan, 2009/2011». (2010). Web:<http://www.ifla.org/files/cataloguing/plans/strategic-plan_2009-2011.pdf>.

—. «Strategic Plan, 2011/2013». (2012). Web: <<http://www.ifla.org/en/node/1959>>.

International Federation of Library Associations and Institutions. Cataloguing Section. Standing Committee. «Minutes. San Juan, Puerto Rico, 2011». (2011). Web: <http://www.ifla.org/files/cataloguing/reports/meeting_2011.pdf>.

Principi di catalogazione internazionali: una piattaforma europea? Considerazioni sull'IME ICC di Francoforte e Buenos Aires. Atti del convegno internazionale, Roma, Bibliocom, 51. Congresso AIB, 27 ottobre 2004. A cura di Mauro Guerrini. Roma: Associazione italiana biblioteche, 2008. Print.

Rius, Assumpció Estivill. «Estado actual de la normativa de catalogación. Primera parte, El escenario internacional». *BiD: textos universitaris de biblioteconomia i documentació* 22. (2009). Web: <<http://www.ub.edu/bid/22/estivill2.htm>>.

Tillett, Barbara B. *The Bibliographic Universe and the New IFLA Cataloging Principles. L'universo bibliografico e i nuovi principi di catalogazione dell'IFLA. Lectio Magistralis in Library Science. Florence University, March 14, 2008.* Fiesole: Casalini libri, 2008. Print.

Tillett, Barbara B., et al. «International Meeting of Experts for an International Cataloguing Code: IME ICC. Report of the 1st Meeting, Frankfurt, Germany, July 28-30, 2003». *International Cataloguing and Bibliographic Control* 33.2. (2004): 39–41. Print.

Watanabe, Takahiro. «Authority Control: Today and the Future». *Journal of information science and technology association (Joho no Kagaku to Gijutsu)* 60.9. (2010): 371–377. Print.

Weih, Jean e Lynne C. Howarth. «Uniform titles from AACR to

RDA». *Cataloging & Classification Quarterly* 46.4. (2008): 362–384. Print. Also available on the web: <http://bing.exp.sis.pitt.edu:8080/webdav/lis2000/Miscellaneous/howarth_3.pdf>.

Willer, Mirna. «Third Edition of UNIMARC Manual: Authorities Format. How Does it Implement Concepts from the FRAD Model and IME ICC Statement of International Cataloguing Principles». *International Cataloguing and Bibliographic Control* 38.4. (2009): 68–70. Print. Also available on the web: <<http://www.ifla.org/files/hq/papers/ifla75/135-willer-en.pdf>>.

Willer, Mirna e Ana Barbaric. «Međunarodna kataložna načela: prikaz i analiza. International cataloguing principles: review and critical analysis». *Vjesnik Bibliotekara Hrvatske* 52.1-4. (2009): 18–62. Print. Also available on the web: <[http://www.hkdrustvo.hr/datoteke/749/vbh/God.52\(2009\),br.1-4](http://www.hkdrustvo.hr/datoteke/749/vbh/God.52(2009),br.1-4)>.

Works cited

- Chaplin, A. H. and Dorothy Anderson, eds. *International Conference on Cataloguing Principles. Paris, 9th-18th October, 1961. Report*. London: Organizing Committee of the International Conference on Cataloguing Principles, 1963.
- Commissione permanente per la revisione delle regole italiane di catalogazione, ed. *Regole italiane di catalogazione. REICAT*. Roma: ICCU, 2009.
- Cutter, Charles A. *Rules for a Dictionary Catalog. 4th edition rewritten*. Washington (D.C.): Government Printing Office, 1904.
- Guerrini, Mauro. "Cataloguing based on bibliographic axiology". *JLIS.it* 1.1. DOI: [10.4403/jlis.it-11](https://doi.org/10.4403/jlis.it-11). (2010).
- IFLA Study Group on the Functional Requirements for Bibliographic Records. *Functional requirements for Bibliographic Records. Final Report*. München: Saur, 1998.
- IFLA Working Group on Functional Requirements and Numbering of Authority Records (FRANAR). *Functional Requirements for Authority Data. A Conceptual Model. Final Report, December 2008*. Ed. Glenn E. Patton. München: Saur, 2009. Traduzione italiana: Requisiti funzionali per i dati di autorità. Un modello concettuale. Relazione finale, dicembre 2008. A cura di Glenn E. Patton. Approvata dagli Standing Committee dell'IFLA Cataloguing Section e dell'IFLA Classification and Indexing Section, marzo 2009. Edizione italiana a cura dell'Istituto centrale per il catalogo unico delle biblioteche italiane e per le informazioni bibliografiche. Roma: ICCU, 2010.
- IFLA Working Group on Principles Underlying Subject Heading Languages. *Principles underlying subject heading languages (SHLs)*. Ed. Maria Inês Lopes and Julianne Beall. München: Saur, 1999.
- ISBD. *International Standard Bibliographic Description. Consolidated ed.* Berlin: De Gruyter Saur, 2011.
- Joint Steering Committee for Development of RDA (JSC), ed. *Resource Description & Access. RDA*. Chicago: American library association; Canadian library association; CILIP, 2010.
- Joint Steering Committee for Revision of AACR. *Anglo-American cataloguing rules*. Ed. Michael Gorman and Paul W. Winkler. 2nd edition, 1988 revision. Ottawa: Canadian Library Association; Library Association; American Library Association, 1988. Traduzione italiana: Regole di catalogazione angloamericane. Redatte sotto la direzione del Joint Steering Committee for revision of AACR. A cura di Michael Gorman and Paul W. Winkler. 2. ed., revisione del 1988. Edizione italiana. A cura di Rossella Dini e Luigi Crocetti. Milano: Editrice Bibliografica, 1997.

- Lubetzky, Seymour. "The Function of the Main Entry in the Alphabetical Catalogue. One Approach". *International Conference on Cataloguing Principles. Paris, 9th-18th October, 1961. Report*. Ed. A. H. Chaplin and Dorothy Anderson. London: Organizing Committee of the International Conference on Cataloguing Principles, 1963. 139-143.
- . *Writings on the Classical Art of Cataloging*. Ed. Elaine Svenonius and Dorothy McGarry. Englewood, Col.: Libraries Unlimited, 2001.
- Svenonius, Elaine. *The Intellectual Foundation of Information Organization*. Cambridge (Mass.); London: MIT, 2000. Traduzione italiana: Il fondamento intellettuale dell'organizzazione dell'informazione. Trad. di Maria Letizia Fabbrini. Introduzione di Mauro Guerrini. Firenze: Le lettere, 2008.
- Verona, Eva. "Literary Unit versus Bibliographical Unit". *Libri* 9.1-4. (1959): 79-104.
- . "The Function of the Main Entry in the Alphabetical Catalogue. A Second Approach". *International Conference on Cataloguing Principles. Paris, 9th-18th October, 1961. Report*. Ed. A. H. Chaplin and Dorothy Anderson. London: Organizing Committee of the International Conference on Cataloguing Principles, 1963. 145-157.

GIULIANO GENETASIO, Casalini Libri.
giuliano.genetasio@casalini.it

Genetasio, G. "The *International Cataloguing Principles* and their future". *JLIS.it* Vol. 3, n. 1 (Giugno/June 2012): 4936-1-4936-17. DOI: [10.4403/jlis.it-4936](https://doi.org/10.4403/jlis.it-4936). Web.

ABSTRACT: The article aims to provide an update on the 2009 Statement of International Cataloguing Principles (ICP) and on the status of work on the Statement by the IFLA Cataloguing Section. The article begins with a summary of the drafting process of the ICP by the IME ICC, International Meeting of Experts on an International Cataloguing Code, focusing in particular on the first meeting (IME ICC1) and on the earlier drafts of the 2009 Statement. It then analyzes both the major innovations and the unsatisfactory aspects of the ICP. Finally, it explains and comments on the recent documents by the IFLA Cataloguing Section relating to the ICP, which express their intention to revise the Statement and to verify the convenience of drawing up an international cataloguing code. The latter intention is considered in detail and criticized by the author in the light of the recent publication of the RDA, Resource Description and Access. The article is complemented by an updated bibliography on the ICP.

KEYWORDS: Cataloguing; ICP; IFLA; International Cataloguing Principles; Library theory

Submitted: 2012-04-14

Accepted: 2012-04-30

Published: 2012-06-01





The catalogue (and the cataloguer): a defenseless entity?

Carlo Revelli

The scholars who study cataloguing rules tend, duly I would say, to apply them inflexibly in order to create a precise information tool, coherent in its structure, unchanged over time, at least for a given span of time, except for later additions. A tool designed to meet all requirements in any environment and at any time. On the contrary, the people who start from the needs of would be users, will admit that the rules are not the pre-eminent motivation behind the catalogue; rather, this rises from the need to meet those needs, therefore the cataloguing praxis will assume an adaptable mode, open to doubts, starting from the choice of the material as well as from the identification of index accesses. The real need to solve the conflict between the convenience of a coherent tool not inconsistent in its nature, and one that may meet requests obviously inconsistent with one another, is a proof confirming the conflict inherent in human events; the library is an integral part of it since it mirrors the changing aspects and needs of its environment. The library, we may note, shows a series of doubts, the need to harmonize different requirements and requests, all, or almost all, legitimate ones. The issues concerning cataloguing reflect this conflict and are one instance of it. First of all we have requirements that change over time, due to social changes, the development of knowledge and technology



innovations. Even if today the prevailing role of new technologies almost obliterates all other existing reasons, the changes have always been with us, and have required new criteria both for gathering information and for making it available. We owe to Osborn the 70 year old claim that the future has in store for each generation of librarians the task of doing anew the work done by its predecessors (Osborn).

The contrast is between the strictness of the rules, crucial to the very existence of the catalogue, as well as to its structure, and the flexibility crucial for a varied spectrum of searches. In this contrast, not to be easily reconciled, it is the librarian's task to supply a working tool designed to overcome it. The needs of single persons cannot call for an exhaustive treatment in the catalogue which speaks with a language for all people. The conflict between rigidly enforced rules and the grounds of research must be solved: not even homogeneous groups of readers, like the ones in specific institutions, i.e. special libraries, can obliterate individual needs. We find the same conflict in all social situations, I would say in all human relationships. The conflict is worse nowadays, just when the global village tends to merge and assimilate different types of resources; yet, it is in the system itself that the specific needs of a single library as well as of the single person, become prominent. In a different field, Carlo Ginzburg reminds us the far from recent trend to "conform to a standard, to a loss of diversity, a uniformity that weakened all individual and national traditions" (Ginzburg, p. 126). Raffaele La Capria claims: "We usually prefer drowning all differences into the slow, lazy stream that standardizes, receives and drags everything" while "our time favours only small differences" (La Capria, p. 1328-1330). We cannot but cite here the philosopher from Senegal Alassane Ndaw: "The basic issue is to live and to make contradictions real" (Ndaw). Is there real contradiction in today's situation witnessing

on one side an ever increasing amount of un-organized information accessible by keywords, on the other side the rigid organization of bibliographic systems? Yet people need to make “informed choices” rather than being left “in a state of darkness” as advocated by Vanda Broughton who teaches Classification and Indexing at the University College in London interviewed for *CILIP Update* (Broughton, p. 21-23). She claims the need for this work in the library, a work wrongfully judged no more needed thanks to technology. Other scholars, like Stalberg e Cronin (Stalberg and Cronin, p. 124-137), admit that assessing the cost/value ratio of Bibliographic Control is not easy because the value of results must be accounted for. Roberto Ventura (p. 152) speaks about a “reconciliation between the theoretical and the pragmatic level”; he examines two separate but equally necessary perspectives, of which I would consider prominent the purpose, the motivation of the catalogue, that is pragmatic, while its theoretical aspects are a consequence. I do not plan starting a talk about the often reiterated question whether or not librarianship is a science since we should first define the term *science*. The conflict between the strictness of theory and use may be solved applying rules in a flexible way, although it might imply lack of uniformity, a consequence to be avoided because of the growing bulk of our information systems. This is a conflict to be found in every field as proved by presence of difficulties in determining the priorities between local on one side and national/international standards on the other side. We can see it in a search for information about art objects in the wide area of libraries, archives and museums and art galleries (Lim and Li Liew, p. 484-498). The rights of users, that can lead to different access points by form or by subject will give see references in the former case and alternative subject headings in the latter case. Further access points will be added to the one created by the application of rules. This solution mitigates the criterion of

uncritical treatment which Francesco Barberi talked of and I find it satisfactory. The point is not about changing the rules, rather, it is about assessing the level of exhaustiveness in the analysis of documents (let's call them resources, if you like) and in the number of access points, in alternative offers. Examples of alternative headings are not lacking – even if heretical ones – for access to names of persons (the prefix *de* for French family names and for some Italian ones for aristocracy titles or Icelandic names) and there are repeated uncertainties most of all for semantic access. The same is true in the case of classification, with its doubts related to interdisciplinary resources. These, although not a novelty, are definitely growing in number presently. An example: subjects for fiction works, useful for young readers and a particular section of users, but not so in the academic field. The presence of the item – copy, exemplar – is confirmed by the latest rules; it is the fourth entity in FRBR Group 1, it is widely dealt with in AACR2 and in REICAT. A library that is part of an information system can not only add access points useful for it, it can also underline particular features, either positive or negative ones, in the copy it owns.

Let's limit our talk to the relationship between the subject catalogue, both card and electronic, and the catalogue of which it is an integral part. Today the unity of the catalogue is more manifest because the basic "entry" is a common element to be retrieved through its diverse sets of access points but also in the past the unavoidable physical separation did not rule out conceiving the catalogue as a unitary body. Thus, I do not advocate a reversion to Cutter and his Dictionary Catalogue with its sequence for author, subject, title and form entries – later blown up into separate sequences – rather, a prosecution in conformity with a different praxis in its historical stages. Today the unity of the catalogue is proved by FRBR and RDA, though they have not faced the semantic aspect yet, and a full

and definitive application of them is still far away.

The cultural evolution due to changes in society, which also affects language mutability, adds to the identification and evaluation of a subject or subjects which may be conditioned by the type of collection, the functions of the library or its users, the cultural level and individual needs of the people searching the catalogue. The catalogue is a fixed entity, in it the need for consistency granting that procedures can be repeated, may show in comparison vague areas which expose points that do not meet the needs recently transpired, which will be revised in the future, make obvious a certain delay. It is also at odds with a series of differences which add to today's multifarious needs, for instance of diachronic variations. Entries created for a past culture are added to the need for the catalogue to try and offer retrieval tools satisfying varied requirements. Again, we find the juxtaposition of the ideal of a global consideration to a set of searches related to a well-defined body – one or more collections – and a change in requirements linked both to the present and to different times in the past. This concerns the relationships with the past but must not make us neglect the future. Antonio Scurati wondered: "How can one tell the present when there is only the present?" (Scurati, p. 295). Talking of retrieving the past with our present day vision Ventura wonders: "What about rediscovering the past apparently buried in the past authors when we find in them elements useful in facing issues hotly debated and make them the object of renewed interest?" (Ventura, p. 126). This retrieval of the past for the benefit of present knowledge is quite true but how will documents reflecting learning at different times relate with each other when they have access under one and the same word? Sure a filter by date usually, not always, may be of help. What is modified is the connotation of a term. The cultural changes from one generation to the next one or within a generation, are attested in an article by

Matt Johnson on the deletion of references to *Sexual perversion* from the terms *Homosexuality* and *Lesbianism*; it anticipating by two years the depathologizing of homosexuality by the American Psychological Association in 1974 (Johnson). Something like that took place in Italy in the same period: the librarian to whom was addressed a complaint about a reference from *Omosessualità* to *Perversione sessuale*, could show to that user the list of corrections that were being done in the subject catalogue, among them there was the deletion of that reference.

In the online catalogue, when we state the order of terms in a string we add to the pre-coordinated research the advantages of the post-coordinated one. The latter is the only one available when we are given isolated terms. When there are various subjects in a document these may muddle the Boolean search – not a favourite one for patrons, as proved by inquiries carried out in several countries. The number of terms forming a subject string depends on the number of concepts making up a subject. The card catalogue and the subject indexes were characterized by a limitation in the number of terms, rather random, but this no longer has a motivation. It is confirmed by Joseph Miller, editor of the largely used *Sears List*, now in its 21st edition, interviewed for *Cataloging & Classification Quarterly*. Miller perceives it as a “good example of the change in thinking” (Rofofsky Marcus). Then we have the ever present issue of language: the choice and organization of terms is an element that is necessary but not sufficient, since availability of access for users implies the use of the same language. What I talk of is not only the language and terminology used in catalogues but the sum of information embedded in them in order to make searching easier. Therefore, we need the catalogue to be so organized as to allow the uncontrolled language of readers to communicate with the controlled language of the catalogue. This guidance to the use of the catalogue must

be part of its structure as we cannot expect our patrons to study cataloguing rules. This is one of the points behind the limited use of the card subject catalogues when there was no guide in them, no definitions, no linking devices, which were to be found only in the subject indexes used by indexers.

A series of doubts that make one wonder whether cataloguing rules are still needed, whether we must acquire materials when we can retrieve information about them otherwise, or recognize the individual peculiarities of a single library, whether to disown the need for a librarian who chooses, organizes and makes available the library documents, who is willing to give advice to a public that no longer asks for it. Some people see in the virtual library the opposite of the physical one, as if physicality were not a prerequisite for a library. Doubts about the profession, its qualifications and staff training, too. These have always been with us accompanying the evolution of human endeavours, today more than in the past for the lapse of time given to a cultural formation is shorter and shorter; this is particularly true for activities in the field of communication and information. As it is often the case, retrieval of what pertains to the past and may be useful to better understand the present follows the excitement of novelty united to a refusal of past experience. If we restrict our talk to the issues on semantic cataloguing, we acknowledge the need to give information in the form of bibliographic access or direct help to users, and this implies admitting that there is still a need for a linking tool not to retrieve a past culture but to conveniently adjust to a new situation. On the point of rigid opinions we may listen to Alberto Salarelli who talks about the “false myth” of disintermediation” (Salarelli).

We get to one of the most controversial issues: the usefulness of the librarian, even of the catalogue, as tools to link users and information. In his guide to RDA, in a way similar to Broughton’s

thesis, Chris Oliver claims “though we should be wise to think deeply about how to adapt to changing social needs” (Oliver). This confirms that evolution does not obliterate the past but it modifies it transforming it. The first chapters are intended to overcome the fear of novelty, as it is proved by a review of Oliver’s work (Rev. of *Introducing RDA: a guide to the basics*).

Joseph Miller, too, comforts us, near the end of the above mentioned interview, that “there is an amount of good libraries, good cataloguing and good serious cataloguers. We must not give up what we all know is of great worth: cataloguing information”. Patrick Bazin, former head of the Municipal Library in Lyons, now head of the Bibliothèque Publique d’Information in Paris, and Ulrich Johannes Schneider, head of the University Library in Leipzig, in a very interesting conversation agreed that the library – as well as the book – will survive, even in the presence of a plurality of information sources, although losing its *Besonderheit*, with the task of classifying and informing only about its own resources “but the cognitive aspect of this is still to be understood” (Bazin and Schneider), as pointed out by Bazin himself. We can see in this case, as in other ones, that the existence of the printed book, of the library, of the catalogue, of the librarian – questioned or declared close to extinction, is valued by the most qualified experts who do not reckon it a surviving entity or a preservation of the past, but a progress that implies and demands full upgrading. There is no doubt that the reiterated reference to the values of the library and the librarian might seem a worrying sign, a sort of lamentation for a set of activities whose need is fading. An American librarian, answering to the economic grounds voiced by those who would reroute the scanty financial investment wondering if there is still a need for the library catalogue, senses “a dramatic change in the profession”. A change with positive results since the new rules in RDA allow us to get over the

“digital tsunami” thanks to a unitary vision of the catalogue and of the resources, with a flexibility apt to treat its “changing landscape”. Authority control and data recording at a lower level would instead depreciate the information service. A greater workload with fewer resources, just when “the need for cataloguers is growing, now more than ever”, because they “are the link in the chain that helps get information where it needs to be” (Cerbo II). On the issue of survival we may recall Robert Darnton’s words about the future of the book.¹ The tasks of the library are still with us, certainly with a less passive attitude than in the past. Its direct action on its patrons renews and elates its functions, in the academic library for its relationship with the teaching activities, as confirmed by the growing popularity of the *Learning Centres*, in the public library for the services offered to a more and more varied public of users who have in common their appreciation of the library as a reference frame for all the citizens.

¹Darnton is well known in Italy and *Il futuro del libro* is the title of the Italian translation, published by Adelphi in 2011 (*The case for books. Past, present, and future*).

Works cited

- Rev. of *Introducing RDA: a guide to the basics*, by Chris Oliver. *Zeitschrift für Bibliothekswesen und Bibliographie* 58.1 (2011): 55.
- Bazin, Patrick and Ulrich Johannes Schneider. "Die gesellschaftliche Aufgabe der Bibliothek". *Zeitschrift für Bibliothekswesen und Bibliographie* 58.1. (2011): 28–33. (Cit. on p. 8).
- Broughton, Vanda. "Making sense of the world with Ranganathan and a fluffy boy". *CILIP Update*. (September 2011). (Cit. on p. 3).
- Cerbo II, Michael A. "Is there a future for library cataloguers?" *Cataloging & Classification Quarterly* 49.4. (2011): 323–327. (Cit. on p. 9).
- Darnton, Robert. *The case for books. Past, present, and future*. New York: Public Affairs, 2011. Italian translation: *Il futuro del libro*. Milano, Adelphi, 2011. (Cit. on p. 9).
- Ginzburg, Carlo. *Il filo e le tracce*. Milano: Feltrinelli, 2006. (Cit. on p. 2).
- Johnson, Matt. "Transgender subject access: history and current practice". *Cataloging & Classification Quarterly* 48.8. (2010): 661–683. (Cit. on p. 6).
- La Capria, Raffaele. "Il sentimento della letteratura". *Opere*. Milano: Mondadori, 2003. 1267–1381. (Cit. on p. 2).
- Lim, Shirley and Chern Li Liew. "Metadata quality and interoperability of GLAM digital images". *Aslib Proceedings* 63.5. (2011). (Cit. on p. 3).
- Ndaw, Alassane. *La pensée africaine. Recherches sur les fondements de la pensée négro-africaine*. Dakar: NEAS, 1997. (Cit. on p. 2).
- Oliver, Chris. *Introducing RDA: a guide to the basics*. Chicago: ALA, 2010. (Cit. on p. 8).
- Osborn, Andrew D. "The crisis in cataloguing". *The Library Quarterly* 11.4. (1941): 393–411. (Cit. on p. 2).
- Rofofsky Marcus, Sara. "An interview with Joseph Miller". *Cataloging & Classification Quarterly* 49.2. (2011): 127–139. (Cit. on p. 6).
- Salarelli, Alberto. Rev. of *Nuovi metodi di gestione dei documenti multimediali*, Milano, *Editrice Bibliografica*, 2010, by Roberto Raieli. *Biblioteche oggi* (Dicembre 2010): 54–55.
- Scurati, Antonio. *Gli anni che non stiamo vivendo*. Milano: Bompiani, 2010. (Cit. on p. 5).
- Stalberg, Erin and Christopher Cronin. "Assessing the cost and value of bibliographic control". *Library Resources and Technical Services* 55.3. (2011). (Cit. on p. 3).
- Ventura, Roberto. *Il senso della biblioteca*. Milano: Bibliografica, 2012. (Cit. on pp. 3, 5).

CARLO REVELLI, già Biblioteche civiche torinesi.
carlorevelli@tiscali.it

Revelli, C. "The catalogue (and the cataloguer): a defenseless entity?". *JLIS.it* Vol. 3, n. 1 (Giugno/June 2012): 5621-1-5621-11. DOI: [10.4403/jlis.it-5621](https://doi.org/10.4403/jlis.it-5621). Web.

ABSTRACT: The author reflects upon the work of cataloguing in present times. He wonders how to harmonize the precision and universality of cataloguing rules and the growing of different cultural needs and information behaviours. Against the widespread ideas on "disintermediation" allowed by information and communication technologies, the author reaffirms the crucial role, although with a quite new background, of librarians and cataloguers as intermediaries between users and documents, between citizens and the information world.

KEYWORDS: Bibliographic universe; Cataloging; Intermediation; Librarians; Library theory

ACKNOWLEDGMENT: The article, already published as "Il catalogo (e il catalogatore), vaso di coccio tra vasi di ferro?", *Biblioteche oggi* 29.10 (2011): 8-12, has been translated from Italian by Maria Letizia Fabbrini, with the collaboration of Grazia Di Bartolomeo.

Submitted: 2012-05-08

Accepted: 2012-05-08

Published: 2012-06-01





Dialogo sopra *Il catalogo per soggetti* di Carlo Revelli

Alberto Cheti

L'esperienza attuale ci permette di riconoscere nel passato potenzialità ancora inespresse o solo affioranti (*Carlo Revelli*)

Cercare di definire i punti per noi caratteristici di una tradizione culturale è cercare di dipingere il nostro autoritratto (*Luigi Crocetti*)

LUIGI – Amico carissimo, è molto tempo che non c'incontriamo. Sono felice di vederti. Ricordo affettuosamente i tempi di lavoro in comune.

GIOVANNI – E' il medesimo stato d'animo, il ricordo del lavoro comune, che mi ha spinto, prima di uscire di biblioteca, a salire le scale e ad affacciarmi nella stanza che una volta era il tuo ufficio. Non pensavo però d'incontrarti. Anch'io sono felice di vederti. Sono passati tanti anni. Ma ho conservato il contatto con te, leggendo i tuoi scritti.

LUIGI – Accomodati. Stavo sfogliando un bel libro, prima che tu arrivassi.

GIOVANNI – Posso sapere di che si tratta?

LUIGI – Dell'edizione anastatica dell'opera di Carlo Revelli (*Il catalogo per soggetti*). Era stata pensata, tempo addietro, per farne omaggio all'autore, in occasione dei suoi ottant'anni, e anche per offrire ai bibliotecari italiani l'opportunità di entrare in possesso di un libro

assente dalle librerie da almeno tre decenni.

GIOVANNI – L’ho vista anch’io, con quella bella veste grigio oro che contraddistingue i volumi della collana e che suppongo voglia indicare un contenuto importante, annunciare una preziosa eredità. Quest’opera ha accompagnato la nostra formazione professionale. Durante le tue lezioni sulla soggettazione, capitava spesso di chiederci, di fronte a un problema: qual è la spiegazione che ne dà Revelli? Che si trattasse, per esempio, di stabilire il soggetto vero del documento, nell’incertezza tra la scienza e il suo oggetto, oppure di decidere la forma corretta dell’intestazione, nell’incertezza se dare la precedenza al nome geografico o al nome comune, *Il catalogo per soggetti* rappresentava per noi un punto di riferimento obbligato... Ora però mi chiedo: quale utilità può avere leggere o rileggere oggi un’opera di quarant’anni fa su un tema così sensibile agli sviluppi teorici e ai cambiamenti tecnologici verificatisi nel campo della catalogazione e dell’organizzazione dei cataloghi? Non ci hai invitato a considerare la biblioteconomia, in quanto disciplina-ombrello, come storica, per necessità funzionali, o, meglio, sempre e solo contemporanea all’età in cui agisce? Non ne consegue che le conoscenze biblioteconomiche debbano servire, in ogni epoca, a creare gli strumenti più efficaci per maneggiare la cultura?

LUIGI – Storica non significa priva di tradizione. Anzi, la sicurezza, l’esattezza, la precisione, che dovrebbero essere lo stile della biblioteca, si misurano sulla conoscenza della tradizione: la biblioteca inevitabilmente si pone come tradizione, non importa se antica o recente, o addirittura contemporanea, poiché sempre di una tradizione si tratta. Sta a noi, al nostro modo di guardare il passato, riconoscerla, distinguendo ciò che del passato serve ai bisogni d’oggi da ciò che, al contrario, non ci suggerisce nulla o ha conseguenze sul presente che giudichiamo inaccettabili.

GIOVANNI – Sono d’accordo. Lo sguardo di cui parli è uno sguardo

rivolto al tempo stesso al passato e alla contemporaneità. E' come se le parole, sotto questo sguardo, fossero messe in tensione tra due polarità, fino a quando non risuonano di accenti nuovi, prima inavvertiti, non assumono un rilievo e una prospettiva particolari. Ci sembra, allora, di essere sulla buona strada per intenderle veramente. Non succede subito, però. Il più delle volte occorre leggere e ascoltare a più riprese, anche a distanza di tempo. Di questa esperienza chiamo *risonanza* il momento percettivo, *riconoscenza* quello cognitivo.

LUIGI – In fondo, si tratta di un'esperienza comune, no? A parte i nomi, un po' singolari, che usi per identificarla. Di un libro ci si appropria a tappe. Leggendo a tappe, o anche leggendo senza sosta, ma poi tornandoci sopra a distanza (per i libri veramente importanti, anche a distanza di anni). Per questo, ho raccomandato ai bibliotecari l'appropriazione direi fisica del libro, la sua costante presenza, la sua perpetua disponibilità, che sole possono garantire quella molteplicità di occasioni, fuori della fretta e della programmazione, necessaria alla totale acquisizione mentale di un'opera.

GIOVANNI – Il messaggio è chiaro. Tuttavia, tornando al tema della tradizione, si deve ammettere che tra il passato e l'oggi sembra essersi determinato uno iato, una frattura insanabile, causata soprattutto dalla rivoluzione informatica.

LUIGI – Negli ambienti d'oggi troneggiano gli elaboratori. Con essi sappiamo di essere (ma forse converrebbe ormai dire essere stati) coinvolti in una vera svolta tecnologica, di dimensioni e intensità non paragonabile alle varie svolte precedenti. Due aspetti ci riguardano più direttamente: le capacità di registrazione e quelle di comunicazione, che sono aumentate a dismisura. Queste hanno conseguenze dirette anche sul lavoro bibliotecario. Cadono regole, divenute in breve tempo obsolete perché "economiche" e non "logiche", si sfasciano un poco i codici di catalogazione, è possibile

concepire in maniera nuova i cataloghi e qualsiasi altro genere d'indice, si è diffuso il fenomeno della catalogazione derivata. . . Tuttavia, la biblioteconomia non risiede nelle regole che ha elaborato per secoli, ma nel fatto che quelle regole derivano dalle possibilità e dai modi di ricerca sviluppati nei millenni dalla specie umana; nella logica del cartografo e in quella del linguista, che significano esattezza e standardizzazione. Come nella cartografia, i suoi dati continuano a pretendere l'esattezza e a voler essere offerti in forma congrua; i nomi e le configurazioni cercate debbono essere cercabili e trovabili secondo sistemi predisposti d'identificazione e di localizzazione. Quanto alla lingua, quella che si parla tra le biblioteche non è la nostra lingua personale, la nostra *parole*; è inane disquisire di linguaggio della biblioteca se ogni biblioteca ne ha poi uno suo proprio. In una disciplina eminentemente formale come la biblioteconomia il linguaggio comprensibile non può nascere che dall'accordo dei parlanti.

GIOVANNI – Dunque, se ho ben compreso, l'appellativo astorico, riferito alla biblioteconomia, comporta quasi un'attenuazione degli effetti che su di essa hanno certi cambiamenti tecnologici.

LUIGI – Sì. Le sue conoscenze seguono sviluppi logici, piuttosto che quantitativi, alcuni acquisiti storicamente, e cioè culturali, altri forse di carattere addirittura biologico e propri della mente umana. Perciò lo iato di cui parli non esiste, oppure si riduce al dover avere confidenza con strumenti prima impensabili.

GIOVANNI – L'abbiamo presa un po' alla larga, ma devo ammettere che questo preambolo mi ha ben disposto nei confronti dell'oggetto del nostro colloquio. Dimmi, allora, quali sono oggi i motivi di interesse di quest'opera di Revelli.

LUIGI – Tralascio d'insistere su quelli storico-bibliografici, su cui possiamo essere tutti d'accordo. Prendi, per esempio, il paragrafo sui rapporti tra catalogo per soggetti e catalogo sistematico e avrai

su questo punto un quadro dello stato della discussione all'epoca, sintetico ma completo, da cui partire per apprezzarne gli sviluppi successivi. Piuttosto, voglio farti notare un dettaglio, piccolo ma significativo: la sorprendente convergenza tra il titolo del saggio introduttivo (Esattezza e praticità) e l'inizio della nuova prefazione autoriale, che dice:

Nella mia esperienza l'interesse nei riguardi della catalogazione per soggetti è legato intimamente con il servizio al pubblico. Non è questa in particolare un'eccezione nel confronto con gli altri lavori in biblioteca, ma i lunghi anni trascorsi a contatto diretto con i lettori mi hanno insegnato a non separare gli aspetti teorici, scientifici se vogliamo usare questa parola, dai risultati pratici derivati dagli strumenti messi a loro disposizione, la cui organizzazione dev'essere tale da giustificare l'esistenza.

Come vedi, entrambi i punti colgono un motivo chiave di quest'opera: unire esattezza e praticità, trovare il giusto equilibrio tra l'interesse teorico e quello pratico, seguire soluzioni logiche senza rinunciare a valutarne l'utilità ai fini della ricerca, adottare regole coerenti senza perdere di vista le esigenze del pubblico. Questo insegnamento è tanto più importante in quanto tutto ciò che dovrebbe rispondere al concetto di servizio da noi non funziona. La società ha cominciato a non comprendere più l'utilità, l'importanza dei grandi depositi del sapere conservati nelle nostre biblioteche, ingestiti e/o ingestibili. La società chiede che siano servizi. Quando vogliamo che accanto ai tesori antichi si accumulino quelli moderni, che gli uni siano posti in relazione con gli altri e insieme interagiscano creando informazione e cultura, bisogna costruire servizi.

GIOVANNI – All'epoca in cui uscì *Il catalogo per soggetti*, gli inizi degli anni Settanta, fu anche una lezione anticipatrice, lungimirante verso la primavera delle biblioteche pubbliche italiane che ancora doveva arrivare. Oggi, per dirla in modo un po' pedissequo, po-

tremmo sintetizzarla così: identificare il soggetto di un documento, esprimerlo mentalmente in un enunciato e rappresentarlo coestensivamente in una stringa, controllare e collegare i termini che ne fanno parte in un vocabolario sono operazioni intermedie di un processo unitario, che ha come obiettivo il recupero ottimale di documenti pertinenti alla richiesta di un utente. I principi e le regole che le governano hanno un duplice compito: assicurare la coerenza interna del processo e, al tempo stesso, concorrere al risultato finale. Prestazioni di processo e prestazioni di servizio, le une in funzione delle altre – direbbe un esperto di management. Dalle nozioni che chiamiamo *uniformità, univocità, esaustività, coestensione, specificità, predittività* scaturiscono, infatti, altrettante direttive operative da cui dipendono i requisiti funzionali e la qualità dell'indicizzazione per soggetto. Occorre però vedere anche l'altra faccia della medaglia: se ai principi e alle regole chiediamo conto della loro efficacia, oltre che della loro coerenza, dalla pratica dobbiamo pretendere che si evitino le scorciatoie, le soluzioni empiriche... Ma ora torniamo al libro di Revelli. Dalla sua prima uscita a oggi non hanno fatto la loro comparsa solo gli elaboratori. Pensa agli standard internazionali, intorno alla metà degli anni Ottanta: da qualche tempo sono già tutti in fase di revisione, per adeguarli a un obiettivo cruciale, l'interoperabilità. E la *Guida GRIS*, il Nuovo soggettoario?

LUIGI – E' vero, ma non sono proprio le successive acquisizioni a far nascere nuove domande da porre a un libro come questo? O pensi che la sua comprensione si sia esaurita con le domande di allora?

GIOVANNI – Quali sono queste nuove domande?

LUIGI – Non ho una risposta bell'e pronta. Posso solo fornirti degli appigli, suggerendoti la lettura di qualche brano del libro, dov'è presumibile trovare indizi utili, più probabili in quei passaggi del testo che si prestano a una duplice interpretazione: ancorata al passato, ossia aderente al contesto culturale, biblioteconomico, organizzativo

e tecnologico in cui è nata l'opera, e proiettata verso il futuro, ossia antcipatrice degli sviluppi successivi della teoria e della pratica dell'indicizzazione per soggetto. Forse, il primo paragrafo del quarto capitolo su *La determinazione del soggetto*, un tema di cui ci siamo occupati a lungo entrambi, può essere un buon inizio... Ne possiamo riparlare tra qualche giorno, se sei d'accordo.

GIOVANNI – Ben volentieri.

LUIGI – Potremmo incontrarci di nuovo qui, mercoledì prossimo, alla stessa ora.

GIOVANNI – Allora, a mercoledì.

* * *

GIOVANNI – Ho letto i tre capitoli centrali del libro di Revelli, dedicati ai problemi tecnici della soggettazione, diciamo, alle tre operazioni fondamentali in cui si articola questa attività. Poi sono tornato all'inizio, proprio al punto che mi avevi segnalato, e ho indugiato su un brano che mi sembra esprimere con chiarezza e lungimiranza i principi che governano l'indicizzazione per soggetto. Non è breve, ma temo che a riassumerlo si corra il rischio di sminuirne la portata.

Perché un catalogo per soggetti risponda alle esigenze per cui è stato formato, occorre in primo luogo che chi vi lavora sia persona in grado di analizzare rapidamente il contenuto di un'opera e di esprimerlo poi in modo sintetico e tale da poter essere reperito facilmente; l'esistenza di un elenco di voci non esime il catalogatore dal compito di individuare l'oggetto dell'opera, per dargli forma concreta attraverso un'intestazione ricavata dal soggettario. A questo si aggiunge che i problemi uguali devono essere sempre risolti allo stesso modo e che i problemi analoghi devono sempre trovare una soluzione analoga, il che significa che ogni intestazione si deve riferire a opere o

parti di opere di contenuto uguale, non assegnabile a nessun'altra intestazione. Significa anche che se a un problema relativo all'intestazione si dà una certa soluzione, è indispensabile per l'uniformità del catalogo comportarsi in modo conseguente in presenza di problemi che, con altri termini, presentino il medesimo meccanismo logico. Questa uniformità di comportamento all'interno del catalogo è determinante per la sua buona riuscita.

Vi sono tre passaggi da sottolineare: a problemi uguali soluzioni uguali, a problemi analoghi soluzioni analoghe, problemi e soluzioni analoghe in presenza del medesimo meccanismo logico.

LUIGI – Che significa? Spiegati meglio.

GIOVANNI – Il primo passaggio cade sotto il campo di applicazione dei principi di uniformità (un soggetto un'intestazione) e di univocità (un'intestazione un soggetto), poiché sono proprio questi principi a garantire l'identità del soggetto e della sua rappresentazione. Ne consegue anche che a opere diverse, ma di contenuto uguale, ossia che hanno lo stesso soggetto, debba essere assegnata la stessa intestazione. L'analogia, invece, ha a che fare con soggetti diversi e, dunque, con intestazioni diverse, che, tuttavia, presentando lo stesso meccanismo logico, devono avere una forma analoga. Si tratta dell'articolazione di quella uniformità di comportamento da cui dipende la buona riuscita del catalogo: uniformità e univocità dell'intestazione, coerenza dell'analisi concettuale dei documenti, predittività delle forme di rappresentazione dei soggetti. Uno stesso requisito applicato a referenti diversi, rispettivamente: la singola unità del linguaggio (termine o stringa), gli enunciati di soggetto, il linguaggio nel suo insieme (vocabolario e sintassi).

LUIGI – Ammetto che il brano letto è una sintesi efficace dei principi a cui deve ispirarsi l'indicizzazione per soggetto. Ma che cosa c'è di particolarmente originale in queste raccomandazioni?

GIOVANNI – Mi ha colpito il richiamo alla necessità di compor-

tarsi in modo conseguente in presenza di problemi che, con altri termini, presentino il medesimo meccanismo logico. Cos'è questo meccanismo logico, su cui si basa l'analogia dei problemi e delle loro soluzioni?

LUIGI – Non vorrei che tu corressi il rischio di partire subito con il piede sbagliato. Non ricordi la critica di Serrai proprio al modo in cui Revelli adopera il termine logica?

GIOVANNI – Sono andato a rileggere le pagine di *Indici logica e linguaggio* a cui ti riferisci. Per la verità, prendono le mosse dal capoverso immediatamente successivo, che dà luogo a una disamina di Serrai sulla natura logica del processo di indicizzazione come un processo di abduzione. Ma affrontare ora questo tema ci porterebbe troppo lontano. Della riflessione di Serrai ho colto, invece, l'invito iniziale a delimitare l'ambito entro il quale il termine logico può acquistare un preciso valore di interpretazione e una netta funzione di chiarificazione.¹ Come interpretare, dunque, il meccanismo logico a cui si riferisce Revelli? E' il meccanismo, mi pare, che sottostà alle regole che anni più tardi saranno poste alla nostra attenzione dagli standard internazionali. Si costruisce la regola su un criterio (caratteristica, proprietà, funzione, relazione) di natura prevalentemente linguistica (grammaticale, semantica, sintattica), applicabile alla generalità dei casi: per esempio, il criterio della *numerabilità* per stabilire la forma plurale o singolare dei termini. Si verifica la rispondenza del singolo caso a quel criterio – ossia, se il termine è numerabile o non numerabile – e si applica a esso la regola. L'insieme dei casi che soddisfano la medesima condizione – per esempio, l'insieme dei termini numerabili – avrà, dunque, una soluzione analoga.

LUIGI – Com'è attestato il funzionamento di questo meccanismo

¹Per la discussione sull'uso di "logica" e "processo logico" in quest'opera di Revelli, v. Serrai (p. 48-53).

nell'opera di Revelli?

GIOVANNI – Be', devo ancora approfondire. Un punto preciso però mi ha dato subito nell'occhio, là dove, nel capitolo su *La forma dell'intestazione*, si analizza la natura delle sottovoci. Revelli individua una condizione generale di ammissibilità delle sottovoci e alcune loro funzioni specifiche. La condizione generale stabilisce che una sottovoce non possa riferirsi a un soggetto solo. Infatti, se così fosse, dovrebbe affrancarsene, diventando essa stessa una voce autonoma e indipendente, senza che con ciò venga compromessa l'integrità del soggetto. Nel caso, per esempio, di *Ruote per automobili*, il termine *ruote* di per sé non si riferisce solo alle *automobili*, potendo indicare anche la parte di altri tipi di macchinari; dunque, è ammissibile come suddivisione di *Automobili*, anzi necessaria, se si vuole distinguere le ruote per automobili dalle ruote in generale o da altri tipi di ruote. Non si può ammettere, invece, *Cuore* come sottovoce di *Apparato circolatorio*, poiché nel significato del primo termine è già implicito il secondo, che perciò non ha bisogno di essere esplicitato nella voce per preservare il senso di questa appartenenza e con esso l'integrità del soggetto. Dunque, il non potersi riferire a un soggetto solo da parte di una sottovoce, secondo l'espressione di Revelli, rappresenta il discrimine tra relazioni esplicite e implicite, tra relazioni sintattiche e semantiche. Revelli però ammette: la constatazione che la sottovoce non può riferirsi a un soggetto solo è un elemento che permette di escludere innumerevoli espressioni dal numero delle sottovoci, ma che non permette di stabilire con certezza se una certa espressione possa essere considerata sottovoce. In altre parole, il criterio consente di stabilire se una certa relazione debba essere indicata esplicitamente nella voce di soggetto, ma non in quale ordine debbano essere citati i termini correlati. Ecco, allora, che vengono in soccorso le funzioni specifiche, le quali identificano, seppure in modo non formalizzato, nozioni riconducibili ai ruoli

sintattici: parti, proprietà, condizioni, attività, aspetti, punti di vista, forme.

Le sottovoci sono sempre espressioni generiche, applicabili come tali a una o più categorie di soggetti o addirittura a tutti i soggetti (ad esempio, *Bibliografia*) e ne possono esprimere una condizione (*Usi e costumi*), una fase della lavorazione (*Pavimentazione*), un aspetto storico (*Politica estera*), un punto di vista (*Concezione marxista*), ecc.

E' riconoscibile qui, in forma embrionale, quel meccanismo logico per stabilire l'ordine di citazione dei termini nella stringa di soggetto che abbiamo conosciuto nel PRECIS, poi adottato nella *Guida GRIS* e applicato nel Nuovo soggettario.²

LUIGI – Alla logica si fa riferimento spesso in queste pagine di Revelli, come antidoto alle soluzioni empiriche. A essa, tuttavia, fa da contrappeso l'uso comune del linguaggio, il cui richiamo serve a evitare arbitrî linguistici, ossia espressioni insolite, inaccettabili, oltre che a preservare il rapporto biunivoco tra termine e concetto. Su questo vorrei leggerti un altro brano del libro:

² L'accostamento tra le funzioni attribuite da Revelli alle sottovoci e i ruoli sintattici del PRECIS e della *Guida GRIS* può risultare, dal modo spiccio e il tono perentorio di Giovanni, un po' troppo ravvicinato. Per ristabilire le giuste distanze e proporzioni, si veda l'inquadramento dei modelli sintattici nell'indicizzazione per soggetto delineato, in modo sintetico ma efficace, da Stefano Tartaglia (Tartaglia, p. 329-333). Per un'analisi delle suddivisioni in termini di ruoli, si rinvia a *Il catalogo* (Revelli e Visintin, p. 283-290) e allo *Studio di fattibilità sul rinnovamento del Soggettario* (Biblioteca nazionale centrale di Firenze, p. 153-186). La conoscenza delle politiche di indicizzazione della Bibliografia nazionale italiana, comprese le scelte sul piano sintattico, in relazione ai caratteri e all'evoluzione della produzione editoriale, ai mutamenti culturali, linguistici e della comunicazione sociale, non può prescindere dalla lettura del volume di Anna Lucarelli (*Produzione editoriale e indicizzazione per soggetto: l'esperienza della Bibliografia nazionale italiana*).

Occorre quindi non solo dare un nome al concetto che già esiste nel nostro intelletto, ma stabilire tra questo nome, o meglio, tra questa espressione e il concetto stesso un rapporto reciproco, in conformità con il quale il concetto non possa essere espresso che con quell'intestazione determinata e quell'intestazione determinata non possa essere la rappresentazione che di quel solo concetto, secondo un'identità accettata dall'uso comune; «il significato d'una parola non dipende dunque dall'uso di un individuo *uti singulus*, ma dall'uso di un individuo in quanto inserito in una certa comunità storica e quindi da ciò obbligato... a evitare ogni arbitrio individuale nell'attribuire un significato a una parola».

L'ultima frase è una citazione dall'*Introduzione alla semantica* di Tullio De Mauro. A prima vista questo principio può apparire di non grande peso, quasi ovvio. Ma che cosa significa in realtà? Significa l'adesione strettissima, completa per quanto è possibile, della soggettazione al sistema linguistico in cui la soggettazione stessa si trova a operare. Nel nostro caso, naturalmente, al sistema linguistico dell'Italia contemporanea. Perché questo? Se vogliamo che *Soggettario* e soggettazione ci restituiscano un'immagine non deformata e deformante della realtà sulla quale operano, in questo caso della nostra cultura, l'unica garanzia che troveremo a questa rispondenza sarà nell'esclusione d'ogni arbitrio di carattere linguistico. Un uso linguistico arbitrario si traduce sempre in un arbitrio scientifico.

GIOVANNI – Chiarissimo. C'è un altro elemento, nel libro di Revelli, che fa da contrappeso alla logica: il giudizio sul punto di accesso più probabile, sull'utilità ai fini della ricerca derivante dall'aver l'uno o l'altro termine in prima posizione nella voce di soggetto. Si tratta, insomma, di una *logica temperata*: dall'uso linguistico e dal criterio dell'accessibilità. Questo secondo elemento presenta però un profilo sfaccettato. Per un verso, il requisito è soddisfatto dall'*uso comune* stesso.

Il bibliotecario dovrà badare, dal canto suo, a fissare le norme più semplici possibili, non in contrasto fra loro, e a stabilire intestazioni accettabili, ammesse dall'uso comune, tali cioè che il lettore che desideri un libro su quell'argomento le cerchi con probabilità nel catalogo.

Per altri versi, invece, costringe a rimettere in discussione il ragionamento logico che dovrebbe portare all'individuazione del soggetto vero e alla scelta della forma corretta dell'intestazione.

LUIGI – Fammi alcuni esempi.

GIOVANNI – *Liguria-Geologia, Milano-Duomo, Diabete-Dietetica, Tumore-Polmone*, ecc.: caso per caso, Revelli fa notare come il pensiero della probabile ricerca da parte dei lettori pesi sulle scelte relative all'individuazione e rappresentazione del soggetto e ci segnala quando le considerazioni pratiche prevalgono sulle ragioni logiche. L'insistenza su questo aspetto, della *praticità*, finisce paradossalmente per rafforzare in noi l'esigenza delle regole più di quanto non emerga dall'illustrazione della loro *esattezza*. Oggi, con la molteplicità e la modularità degli accessi rese possibili dall'informatica, il mix di esattezza e praticità si compone in altro modo... Ma questo è già detto nel saggio introduttivo e non importa insistervi qui... Si è fatto tardi, ora devo andare... Perché non ci vediamo ancora una volta? Hai qualche altro suggerimento di lettura da darmi?

LUIGI – Sarei interessato a conoscere la tua opinione sul brano iniziale del secondo paragrafo del quarto capitolo... Allora a mercoledì prossimo, stessa ora.

* * *

GIOVANNI – Scusa il ritardo. E' molto che aspetti? Sono sempre di corsa... Però, ho svolto il compito che mi avevi dato.

LUIGI – Non preoccuparti. Accomodati. Vogliamo rileggere insieme il brano su cui c'eravamo lasciati la volta scorsa?

In questo senso il principio della scheda unica, in sé non peculiare del catalogo alfabetico, appare come una sua esigenza particolare: «Un solo argomento non può essere rappresentato che da un solo soggetto», proposizione riaffermata più recentemente da Harry Dewey: «solo un'intestazione di soggetto rappresenterà ogni singolo argomento, non importa in quali diversi modi questo concetto possa essere descritto nei titoli dei libri». Per questo motivo è suggeribile che alle opere unitarie sia sempre assegnata una sola intestazione per soggetto e che gli eventuali altri punti di vista si ricavino dalla rete dei richiami, che collegano le voci, e non le opere, tra di loro. Se poi esistono punti del libro particolarmente sviluppati dall'autore, questi potranno dar luogo ad altre intestazioni, che ovviamente non si riferiranno all'insieme dell'opera.

In che senso, secondo te, il principio della scheda unica è un'esigenza particolare del catalogo per soggetti?

GIOVANNI – In effetti, per comprenderlo occorre partire dalla fine del paragrafo precedente: compito primo del catalogo per soggetti è quello di indicare direttamente le opere che trattano dell'argomento desiderato, senza doverlo inscrivere in una cerchia più ampia di argomenti sovraordinati o coordinati. E' questo contenuto specifico dell'opera a essere rappresentato nel catalogo con una scheda unica, quando l'opera è unitaria. Un solo argomento un solo soggetto e un'unica scheda. La definizione, citata nel brano appena letto, è di Enrico Jahier, millenovecentotrentotto.³ Il ragionamento di Revelli si conclude poi con la raccomandazione che eventuali altri punti di vista si ricavino dalla rete dei richiami, che collegano le voci, e non le opere, tra di loro. E' chiaro che questi punti di vista sono quelli offerti dalle relazioni semantiche del vocabolario, che sono appunto legami tra voci. Ne consegue l'invito di Revelli a non creare, per una

³È la seconda delle cinque "regole di assoggettazione" presentate da Jahier al Congresso nazionale dei bibliotecari italiani (Jahier, p. 285).

stessa opera, più intestazioni tra le quali intercorrano rapporti di implicazione logica, gli stessi sui quali è fondata la struttura sindetica dei cataloghi per soggetto. Si può, dunque, sintetizzare così la sua raccomandazione: un'opera unitaria ha un soggetto unitario che dev'essere espresso da un'intestazione unica, alla quale, mediante la struttura sindetica, sono collegate logicamente altre intestazioni, che non hanno bisogno perciò di essere assegnate direttamente all'opera. LUIGI – Intravedo, alla fine del ragionamento di Revelli, la stringa unica raccomandata dal GRIS e, implicitamente, la distinzione tra vocabolario e sintassi, seppure priva di tutte quelle implicazioni che gli standard internazionali e la *Guida GRIS* renderanno esplicite.

GIOVANNI – Sì, ma manca ancora l'ultima frase: se poi esistono punti del libro particolarmente sviluppati dall'autore, questi potranno dar luogo ad altre intestazioni, che ovviamente non si riferiranno all'insieme dell'opera. Completiamo il ragionamento precedente: un'opera unitaria ha un soggetto unitario che dev'essere espresso da un'intestazione unica, la quale si riferisce al tema complessivo, mentre eventuali altre intestazioni potranno essere impiegate per esprimere temi particolari sviluppati nell'opera.

LUIGI – Questa intestazione unica, che esprime il soggetto complessivo dell'opera, è la stringa unica coestesa, no?

GIOVANNI – Suppongo di sì. Se eventuali altre intestazioni da assegnare a un'opera sono ammesse solo in quanto si riferiscono a temi particolari, ne consegue che l'onere di esprimere il tema complessivo – quello che il GRIS definisce *tema di base* – debba essere interamente affidato a un'unica intestazione, evitando la scomposizione di un soggetto unitario in soggetti più semplici, ciascuno dei quali rappresentato da una propria intestazione indipendente. Un'opera unitaria un soggetto unitario e un'intestazione unica, coestesa con il tema complessivo dell'opera e collegata logicamente ad altre intestazioni mediante la struttura sindetica. Quanto si è discostata da queste

raccomandazioni la pratica tradizionale della soggettazione? Non troppo, credo. Se non ricordo male, dall'indagine preparata per lo *Studio di fattibilità*, su un fascicolo della *Bibliografia nazionale italiana*, le schede bibliografiche con più di una voce di soggetto, dovuta ai motivi che si sono detti, risultavano complessivamente poco più del venti per cento del campione analizzato; tant'è che l'indagine, in conclusione, ammetteva che la moltiplicazione delle voci di soggetto non fosse la regola, bensì un rimedio pratico alla limitata efficienza del linguaggio nell'esprimere le relazioni semantiche e sintattiche tra i concetti.

LUIGI – Se si mettono una dietro l'altra la definizione di Jahier, la raccomandazione di Revelli e la pratica della *Bibliografia nazionale italiana*, verrebbe da concludere che la stringa unica coestesa abbia fondamenta ben più antiche, sebbene la sua *esattezza e praticità* debbano essere attribuite senza dubbio al merito del GRIS. . . A questo proposito, mi vengono in mente le parole di Alberto Petrucciani di presentazione della *Guida GRIS*. Aspetta. . . prendo il libro. . . ecco: "un metodo coerente di indicizzazione alfabetica per soggetto non fa che realizzare le aspirazioni che si trovavano alla base dei sistemi tradizionali, ma in maniera più efficiente e più economica, idonea ad essere insegnata e appresa invece che imitata a orecchio, suscettibile di essere applicata in modo decisamente più uniforme in luoghi, tempi e da persone diverse. Potremmo scoprire, allora, che il metodo di indicizzazione che questa *Guida* propone parla quasi sempre un linguaggio che conosciamo, spesso proprio il medesimo, perché è il linguaggio delle cose e dei libri, talvolta un linguaggio più preciso, più articolato, che vuole offrire a chi lo cura e a chi lo legge un filo più lucente e più sicuro per non smarrirsi in un labirinto informativo tanto più vasto e più intricato". . .⁴ E con queste belle parole è venu-

⁴La citazione di Alberto Petrucciani è tratta dalla sua introduzione alla *Guida GRIS* (*Guida all'indicizzazione per soggetto*).

to, forse, il momento di congedarci dalla nostra conversazione. . .

GIOVANNI – Riconosco che esse ne sarebbero la degna conclusione. Non avevamo iniziato proprio parlando di tradizione? Se però hai ancora un po' di tempo, vorrei aggiungere qualcosa. Sarò breve.

LUIGI – Ho tutto il tempo che vuoi.

GIOVANNI – Spiegando, la volta scorsa, il significato di meccanismo logico, ho portato come esempio la regola per l'uso del singolare e del plurale dei termini. Così, sono andato a rileggere, nel quinto capitolo, il paragrafo *Singolare o plurale?*, sebbene si tratti di una questione di importanza minore, come osserva lo stesso Revelli. Inizio da questa citazione:

In altre parole, mentre la forma singolare è adatta in ogni caso a un vocabolario, dove si trovano definizioni di concetti, in un catalogo per soggetti essa non appare sempre opportuna, perché si riferisce all'insieme degli individui che a quella definizione rispondono. La distinzione appare ancora più evidente se si considerano certe suddivisioni delle voci in esame, in particolare le limitazioni geografiche.

Con questa affermazione si esclude l'uso di un'unica forma (singolare) per tutti i soggetti, come avviene in un vocabolario. Perché? Perché la forma singolare, in un vocabolario, indica l'insieme degli individui a cui si riferisce la definizione del lemma; mentre, in un catalogo, ciascun termine che compone un'intestazione può rappresentare il concetto indicandone tutti gli individui, oppure riferirsi solo ad alcuni di essi o a un unico individuo.

LUIGI – Fammi qualche esempio.

GIOVANNI – Questa diversificazione dei termini, riguardo al numero, dipendente dal contesto dell'enunciato di soggetto e della stringa corrispondente, è esemplificata da Revelli con le delimitazioni geografiche e cronologiche: per esempio, *Partiti-Belgio-Sec. XIX* (non *Partito*), *Abbazie-Italia* (non *Abbazia*). Seconda citazione:

La forma dev'essere decisa di volta in volta, ma si badi a non cedere alla prima impressione. A un trattato di limnologia, ad esempio, potrebbe essere assegnata l'intestazione al singolare, *Lago*, se non si pensasse a possibili libri futuri che trattassero dei laghi del mondo o di una regione particolare.

Questa affermazione esclude la possibilità di usare, per uno stesso termine, ora il singolare ora il plurale, a seconda del soggetto, ossia a seconda che ci si riferisca all'insieme degli individui o a individui determinati di una classe: per esempio, *Lago* per un trattato di limnologia e *Laghi* per i laghi italiani. Con l'eccezione del singolare, che sarà sempre da preferire, nell'intestazione, quando ci si riferisce a un'entità individuale:

...la forma plurale *Vertebrati* e *Partiti* esprime meglio un concetto collettivo che al singolare sarebbe più indicato solo per esprimere l'appartenenza a quella categoria da parte di un solo individuo.

Ci vuole, dunque, un criterio predefinito, che consenta di stabilire il numero del termine a prescindere dalla sua occorrenza in un contesto particolare.

LUIGI – Qual è questo criterio?

GIOVANNI – Revelli lo definisce così:

Diremmo piuttosto che si può adottare la forma singolare per quei soggetti che sono pensabili al singolare anche quando siano limitati cronologicamente o geograficamente. Per un'opera sui vini francesi possiamo immaginare l'intestazione *Vino-Francia*, ma per un'opera sui laghi italiani non possiamo immaginare che per scherzo l'intestazione *Lago-Italia*. Non rimane in questo caso che adottare la forma plurale.

È implicito in questo ragionamento il criterio della *numerabilità*, che la *Guida GRIS* pone a base della scelta tra singolare e plurale: si

usa il plurale per i termini che rappresentano concetti numerabili, cioè che designano entità enumerabili (che si possono contare); si usa il singolare per i termini che rappresentano concetti di massa, cioè che designano entità non enumerabili. Ai concetti numerabili corrispondono, nella formulazione di Revelli, i concetti che non sono pensabili al singolare quando siano limitati geograficamente e/o cronologicamente; mentre, ai concetti di massa quelli pensabili al singolare anche in presenza di tali limitazioni. Infatti, la condizione di non essere pensabile al singolare, quando una limitazione spaziale o temporale ne restringe l'estensione, è indice della numerabilità del concetto. Ci può soccorrere, nella scelta, l'analisi categoriale dei concetti?

Il singolare è preferibile per tutti i soggetti che esprimono un concetto astratto; nel caso dei concetti concreti sarà ugualmente da preferirsi, quando il concetto della categoria superi la considerazione dei suoi singoli componenti e la differenza tra le sue varietà.

Seppure in forma generica ed embrionale, vi si intravede quel criterio del *numero della categoria* che nella *Guida GRIS* sarà illustrato in modo preciso e articolato. Insomma, nel testo di Revelli su singolare o plurale sembra di poter leggere, in filigrana, lo stesso meccanismo logico che sottostà alle regole corrispondenti della *Guida GRIS*, sebbene grande sia la loro differenza in termini di esattezza e praticità... È l'ora di salutarci. Non vorrei abusare ancora della tua pazienza. Ho però il presentimento che c'incontreremo di nuovo, in nome dell'amicizia e del lavoro comune di un tempo... E avremo altre occasioni per definire i punti per noi caratteristici di una tradizione... A presto.

LUIGI – A presto, amico carissimo.

Nota

La prima parte della conversazione prende spunto da diversi scritti di Luigi Crocetti («Noi e gli altri: un confronto possibile? Conversazione fra due amici sull'orgoglio di ieri e i complessi di oggi», p. 8-10; «La tradizione culturale italiana del Novecento», p. 76-81; «Lo stile della biblioteca», p. 223-228; poi in: *Il nuovo in biblioteca e altri scritti*, p. 35-40; «Il libro è mio», p. 5; poi in: *Il nuovo in biblioteca e altri scritti* p. 131-133; «Bibliothecarius technologicus: rivoluzione quantitativa o nascita di una nuova specie?», p. 6-10; poi in: *Bibliotecario nel 2000: come cambia la professione nell'era digitale*, p.19-27).

Eccetto qualche espressione di collegamento e alcune brevi annotazioni – il giudizio sull'interesse storico-bibliografico dell'opera di Revelli, l'accostamento tra il titolo del saggio introduttivo e l'inizio della prefazione autoriale, la battuta finale della prima parte della conversazione – le parole pronunciate da Luigi (o attribuite a lui da Giovanni) riproducono quasi letteralmente passi tratti da questi scritti. Ciò non attenua, tuttavia, l'arbitrarietà della selezione e la discrezionalità dell'interpretazione.

Della seconda parte del dialogo segnalo solo il punto in cui Luigi spiega il principio dell'uso linguistico: le parole che seguono la citazione di Revelli sono tratte dalla lezione di Crocetti *Soggettazione* (Firenze, 197?, dattiloscritto), della quale ho dato ampiamente conto nel saggio sulla lezione di Casamassima (Cheti, p. 53-86).

Nella terza parte non figurano citazioni di scritti di Crocetti.

L'indagine, a cui si fa riferimento nella conversazione, sul numero di voci di soggetto per registrazione bibliografica e sui motivi di questa attribuzione da parte della Bibliografia nazionale italiana, si trova nello Studio di fattibilità sul rinnovamento del Soggettario (Biblioteca nazionale centrale di Firenze, p. 139-152).

Riferimenti bibliografici

- Biblioteca nazionale centrale di Firenze. *Per un nuovo Soggettario: studio di fattibilità sul rinnovamento del Soggettario per i cataloghi delle biblioteche italiane*. Milano: Editrice Bibliografica, 2002. 153–186. (Cit. alle pp. 11, 20).
- Cheti, Alberto. «La lezione di Emanuele Casamassima sulla soggettazione». *Piccoli scritti di biblioteconomia per Luigi Crocetti (10 marzo 2007-10 marzo 2008)*. Manziana (Roma): Vecchiarelli Editore, 2008. 53–86. (Cit. a p. 20).
- Crocetti, Luigi. *Bibliotecario nel 2000: come cambia la professione nell'era digitale*. A cura di Ornella Foglieni. Milano: Editrice Bibliografica, 1999. 19–27. (Cit. a p. 20).
- . «Bibliothecarius technologicus: rivoluzione quantitativa o nascita di una nuova specie?» *Biblioteche oggi* 16.4. (1998): 6–10. (Cit. a p. 20).
- . «Il libro è mio». *Bit* 9.35. (1992). (Cit. a p. 20).
- . *Il nuovo in biblioteca e altri scritti*. Roma: Associazione italiana biblioteche, 1994. (Cit. a p. 20).
- . «La tradizione culturale italiana del Novecento». *L'automazione delle biblioteche nel Veneto: tra gli anni '90 e il nuovo millennio, Venezia, Palazzo Querini Stampalia, 29-30 gennaio 1999*. A cura di Chiara Rabitti. Venezia: Fondazione scientifica Querini Stampalia, 2000. 76–81. (Cit. a p. 20).
- . «Lo stile della biblioteca». *La cultura della biblioteca: gli strumenti, i luoghi, le tendenze. Atti del convegno di Biblioteche oggi, Châtillon, 22-24 maggio 1987*. A cura di Massimo Belotti. Milano: Editrice Bibliografica, 1988. 223–228. (Cit. a p. 20).
- . «Noi e gli altri: un confronto possibile? Conversazione fra due amici sull'orgoglio di ieri e i complessi di oggi». *Biblioteche oggi* 11.11. (1993): 8–10. (Cit. a p. 20).
- Gruppo di ricerca sull'indicizzazione per soggetto. *Guida all'indicizzazione per soggetto*. rist. con correzioni. Roma: Associazione italiana biblioteche, 2001. (Cit. a p. 16).
- Jahier, Enrico. «Catalogo a soggetto e schedatura centrale». *Accademie e biblioteche d'Italia* 12.3-6. (1938): 281–292. (Cit. a p. 14).
- Lucarelli, Anna. *Produzione editoriale e indicizzazione per soggetto: l'esperienza della Bibliografia nazionale italiana*. Milano: Editrice Bibliografica, 1998. (Cit. a p. 11).
- Revelli, Carlo. *Il catalogo per soggetti*. Firenze: Le Lettere, 2011. (Cit. a p. 1).
- Revelli, Carlo e Giulia Visintin. *Il catalogo*. Milano: Editrice Bibliografica, 1996. 283–290. (Cit. a p. 11).
- Serrai, Alfredo. *Indici logica e linguaggio: problemi di catalogazione semantica*. Roma: Consiglio nazionale delle ricerche, 1974. 48–53. (Cit. a p. 9).
- Tartaglia, Stefano. «La sintassi delle stringhe di soggetto». *Bollettino AIB* 37.3. (1977): 329–333. (Cit. a p. 11).

ALBERTO CHETI, Comune di Fucecchio.

chetialberto@infinito.it

Cheti, A. "Dialogo sopra *Il catalogo per soggetti* di Carlo Revelli". *JLIS.it* Vol. 3, n. 1 (Giugno/June 2012): 4816-1–4816-22. DOI: [10.4403/jlis.it-4816](https://doi.org/10.4403/jlis.it-4816). Web.

ABSTRACT: Il contributo presenta, in forma dialogica, i contenuti de "Il catalogo per autori" di Carlo Revelli (1970), ripubblicato di recente. Un approccio sistematico alla catalogazione per soggetti, nel quale l'autore unisce, a una notevole esperienza pratica, un'ampia informazione teorica e storica. L'importanza dell'opera nel panorama della biblioteconomia italiana è indubbia, sia come strumento di lavoro in biblioteca e come testo di riferimento per la formazione di bibliotecari. Tutto ciò ne fa un classico della biblioteconomia italiana nel campo della catalogazione.

KEYWORDS: Catalogazione; Catalogo per soggetti; Indicizzazione per soggetto; Revelli, Carlo

ACKNOWLEDGMENT: Il dialogo è una rielaborazione dell'intervento di presentazione dell'edizione anastatica del libro di Carlo Revelli *Il catalogo per soggetti*, tenuto presso l'Università degli Studi di Firenze l'8 novembre 2011, nell'ambito del Seminario di Biblioteconomia diretto da Mauro Guerrini. A Guerrini e a Luigi Crocetti si deve l'iniziativa della pubblicazione di questo libro di Revelli nella collana Pinakes dell'editore Le Lettere

Submission: 2012-04-01

Pubblicazione: 2012-06-01





Ri/leggere Revelli.

Metonimia per una presentazione de *Il catalogo per soggetti* di Carlo Revelli

Pino Buizza

Ritorna in edizione anastatica *Il catalogo per soggetti* di Carlo Revelli (edizione originale Roma: Bizzarri, 1970), con una prefazione dell'autore e un saggio introduttivo di Alberto Cheti. L'opera è ben nota ai catalogatori degli ultimi decenni del secolo scorso, credo poco a quelli dell'ultima generazione. L'uscita nella collezione Pinakes la conferma al rango di classico, un bel riconoscimento, considerando che non appartiene ai vecchi "testi sacri" delle generazioni precedenti. Per chi non studiò la prima edizione, andata ben presto esaurita ma presente in diverse biblioteche italiane ed estere, ricordo, molto semplicemente, che tratta la materia in modo organico e completo.

Una introduzione storica conduce dalle prime forme di registrazione dei libri per argomento, variamente fondate sulle materie o su parole significative del titolo, attraverso passaggi non lineari, fra cui l'introduzione di voci per argomento nei cataloghi per autore, fino al catalogo dizionario e poi alla piena affermazione autonoma del moderno catalogo per soggetti, alla fine del XIX secolo in America e quindi in Europa. Alcuni paragrafi che illustrano la situazione dei cataloghi per soggetti in vari paesi completano il quadro di riferimento e danno precisa dimostrazione della ricchezza di do-



cumentazione e dell'attento studio che stanno alla base dell'opera. Nei successivi capitoli centrali, dopo l'esposizione delle funzioni del catalogo per soggetto, il primo posto spetta alla selezione del materiale e alla politica del catalogo, che va visto in correlazione agli altri strumenti disponibili e privilegiando l'utilità rispetto alla completezza, suggerendo cioè criteri di selezione ed esclusione del materiale da sottoporre. Il soggettario è presentato nella sua struttura e articolazione, come supporto indispensabile alla coerenza interna e diacronica e come strumento di uniformità interbibliotecaria, con l'opzione per un soggettario nazionale a preferenza dell'adozione di uno straniero. Con la "determinazione del soggetto" si raccomanda l'identificazione dell'argomento dell'opera (senza addentrarsi nei problemi dell'analisi concettuale, che rischiano di diventare molto complessi o astrattamente teorici) e del concetto che vi è preminente, il concetto chiave, da rappresentare verbalmente in modo univoco e specifico, con una particolare attenzione alla probabilità di ricerca da parte del lettore: dobbiamo ricordare che al tempo dei cataloghi cartacei la scelta della voce principale risultava determinante per la reperibilità del soggetto nella sequenza alfabetica. Per la forma dell'intestazione sono esaminate tutte le problematiche relative alle sottovoci, ordine di successione incluso, in particolare per i termini geografici. A completare l'insieme delle operazioni, sono esaminate la struttura e la forma del catalogo, sottolineando l'importanza dei collegamenti fra le voci e delle schede guida, e analizzando le diverse possibilità di ordinamento delle schede. Un capitolo finale è dedicato alle biblioteche speciali.

Una caratteristica essenziale del testo è la doppia dimensione su cui il discorso si svolge. Ogni tema, aspetto o problema è affrontato a partire dai dati delle situazioni concrete e secondo un punto di vista tecnico: quale sia, editorialmente e culturalmente, la letteratura da trattare, quali le caratteristiche d'insieme della raccolta per cui si

opera, quali gli interrogativi e le esigenze di ricerca dei lettori, quali le ipotesi concettuali e linguistiche per farli incontrare nelle voci d'indice, singolarmente e nell'insieme delle raccolte. All'esposizione svolta secondo i capitoli citati sopra e di per sé sufficiente a far capire che cos'è, come funziona, si crea e si usa un catalogo per soggetti, è costantemente affiancato un ricco riferimento a opinioni e pratiche espresse e attuate da una grande varietà di bibliotecari e biblioteche che hanno affrontato il tema nel dettaglio dei suoi singoli aspetti. E' impossibile, dopo una lettura così, non rendersi conto esattamente delle dimensioni e complessità dell'attività di soggettazione, anche di singole questioni che a prima vista parrebbero facili o di poco peso. Ci si può esercitare nel consentire o dissentire, in toto o in parte, dalle varie posizioni riportate. Si può ricostruire, un poco per volta, la fisionomia delle diverse scuole e delle linee di evoluzione tracciate da Revelli nel primo capitolo. Se ne ricava in ogni caso un insegnamento di metodo, che nulla dà per acquisito e tutto sottopone a confronto e verifica. Non è difficile poi scorgere qui già pienamente espressa la vocazione di Revelli esploratore, curioso collezionista di tutto quanto si fa, si dice e si scrive nelle biblioteche di ogni parte del mondo, come lo conosciamo attraverso l'apprezzata e inossidabile rubrica *Osservatorio internazionale*, da lui curata su "Biblioteche oggi" da quasi vent'anni, dove i reperti così raccolti e selezionati sono organizzati intorno a nuclei tematici omogenei e offerti a beneficio dell'aggiornamento professionale.

Elogiata dalle numerose recensioni del tempo in Italia e all'estero per accuratezza, vastità di documentazione, equilibrio di pratica, teoria e storia, l'opera divenne l'introduzione italiana alla soggettazione e un importante ausilio per affrontare le numerose problematiche che le scarse indicazioni del *Soggettario* del 1956 lasciavano aperte. Troppo densa e possibilista non divenne quello che non era: un manuale; ma chi si fosse trovato davanti a un dilemma, lì ne avrebbe

trovato l'esposizione analitica sotto tutti gli aspetti, con un ventaglio di soluzioni differenti, proposte in tempi e luoghi diversi, con i pro e i contro di ciascuna, e l'indicazione di quale risultasse alla fine preferibile e perché. Negli anni successivi vennero i manuali, corredati da esemplificazioni commentate e da esercizi, vennero i corsi di formazione, molti tenuti dallo stesso Revelli (personalmente fruii di una tre giorni a otto ore – gennaio 1982 – alla velocità di una 24 ore di Le Mans, ma con estrema chiarezza e precisione, consequenzialità nello svolgimento dei temi, esempi puntuali a corredo, e ne portai via pagine fitte di appunti preziosissimi per gli anni della professione); *Il catalogo per soggetti* scivolava un po' in ombra anche perché esaurito, ma rimaneva il riferimento principe. Venne, come vera novità nella tradizione italiana, il lavoro del Gruppo di ricerca sull'indicizzazione per soggetto dell'AIB, teso ad esplorare possibilità diverse di soggettazione, fondate teoricamente e più funzionali, e portò alla *Guida GRIS*. Giunse poi l'impegno della Biblioteca nazionale centrale di Firenze per il rinnovamento del *Soggettario*, fino alla realizzazione del *Nuovo soggettario* di cui oggi disponiamo. Il quadro italiano è quindi molto modificato. Il riferimento agli standard ISO per il lessico, organizzato in forma tesaurale, comporta l'esplicitazione completa delle relazioni semantiche fra i termini, rispetto alla più semplice struttura sindetica a due relazioni (vedi, vedi anche). La categorizzazione piuttosto rigida delle voci e sottovoci, la povera sintassi di costruzione delle stringhe sono ora rimpiazzate dalle categorizzazioni semantiche del thesaurus e dall'analisi dei ruoli con i criteri sintattici per l'ordine di citazione dei termini.

Testo reso obsoleto, dunque? di puro valore storico? Non si direbbe proprio. Se per un verso non risulta ora di utilità immediata per catalogare secondo il *Nuovo soggettario*, le capillari osservazioni, il continuo dibattito a più voci fra scelte e opzioni argomentate ne fanno un ottimo avviamento all'assunzione mentale della logica e

della tecnica di soggettazione, che è la dote necessaria per operare, qualsiasi sia il linguaggio da adottare in una situazione data, e tanto più nei casi di passaggio da un linguaggio ad un altro. Contro l'insufficienza di un apprendimento puramente applicativo, magari ristretto alla conferma di correttezza, il confronto ragionato fra alternative offre l'allenamento e la strumentazione di base per affrontare consapevoli ed efficaci qualsiasi situazione operativa diversa. Il valore storico inoltre non è la testimonianza di un passato in dissoluzione: i nostri cataloghi per soggetti sono ancora in massima parte cataloghi tradizionali, qualunque sia il rapporto instaurato tra vecchio e nuovo sistema: di cesura fra due tronconi, di convivenza con elementi di distinzione e di collegamento, o qualche forma di traduzione più o meno parziale. La memoria di come il catalogo tradizionale sia stato costruito resta dunque di estrema importanza per il suo uso e la sua decodifica, che si tratti di servizio diretto alle richieste dei lettori o di revisione interna per svecchiare o per ridurre l'attrito col nuovo sistema. Aggiungerei un'osservazione sulla presenza pervasiva nell'opera del dibattito fra catalogo sistematico e alfabetico, allora molto sentito, oggi poco. Se il peso dedicatogli sembra ora eccessivo, ciò non si deve alla avvenuta soluzione del problema, ma piuttosto alla sua dissoluzione nella "facilità" delle ricerche dirette, dove scompare ogni controllo semantico e struttura classificatoria: la sottolineatura risulta ancora un valido richiamo a realizzare coi nuovi dispositivi funzionalità correlate e combinate di ricerca semantica.

Ma la conferma migliore del persistente valore credo venga dal saggio introduttivo di Alberto Cheti (*Esattezza e praticità. Leggere Il catalogo per soggetti di Carlo Revelli*). Avrebbe potuto contenere un autorevole confronto fra il sistema tradizionale e il nuovo, un bel prontuario con tabella che ci risparmiasse di analizzare in proprio e per ogni aspetto ciò che di diverso e di progressivo comporti

un'impostazione rinnovata. Invece, con scelta originale e del tutto apprezzabile, Cheti ci trasporta nel bel mezzo delle lezioni di soggettazione di Luigi Crocetti che utilizzava, dieci anni dopo l'uscita, l'opera di Revelli, e possiamo innanzi tutto coglierne l'importanza nell'atmosfera del tempo, ma scopriamo anche, attraverso acute analisi nel merito di alcuni passaggi del testo (su disciplina vs. oggetto di studio, limitazioni geografiche vs. sottovoci di toponimi, significatività ...) inaspettate consonanze di fondo con i principi e i criteri che in seguito avrebbero prodotto l'elaborazione analitico-sintetica, per quanto diversi risultino la terminologia e gli esiti tecnici. Quarant'anni hanno portato più evoluzione che rivoluzione, come lo stesso Revelli ricorda in prefazione all'anastatica, sottolineando come il cambiamento del catalogo abbia origine e giustificazione in quello di due elementi ad esso essenziali, il pubblico e gli strumenti della comunicazione. Lungo questo filo si potrebbe rileggere l'arco di tempo che separa le due edizioni. Costantemente Revelli ha partecipato all'invecchiamento del *Soggettario* con contributi molto puntuali nel segnalare analiticamente disfunzioni, incoerenze, lacune e proporre correzioni, sempre mantenendosi all'interno dell'impostazione di fondo tradizionale, ma riconoscendo la crescente differenziazione del pubblico e il progressivo impatto dell'automazione e delle nuove tecnologie applicate ai cataloghi come fattori che richiedono rinnovamento, di aprire, per esempio, a forme di soggettazione della letteratura, specialmente per il pubblico giovanile, o di prendere atto della diminuita valenza della prima posizione nella stringa dal momento che ogni termine è ricercabile (e forse in proposito ci si sarebbe aspettati una più forte difesa della stringa ben formata anche in ambiente elettronico, per i valori di significatività, completezza e ordinamento, se la ricerca del termine a prescindere dalla sua posizione è una possibilità in più, ma non in sostituzione della ricerca su liste di stringhe precoordinate).

La cura premurosa per le sorti del vecchio sistema non ha impedito a Revelli di partecipare all'elaborazione del *Nuovo soggettario* (e non si è perduta una competenza unica, di un "maestro", secondo il ringraziamento tributatogli nella *Guida*) e l'ha fatto senza negarsi osservazioni critiche, in particolare contro il rischio che in nome di regole in sé corrette si manchi la corrispondenza con l'esigenza di interrogazione del lettore/ricercatore, e siamo di nuovo al problema della scelta del concetto chiave, in particolare per i soggetti personali e per i toponimi. Il respiro internazionale del lavoro di Revelli induce a richiamare pure alcune linee di sviluppo del catalogo per soggetti all'estero. Da una parte la scuola britannica portò all'elaborazione del PRECIS, adottato dalla *British national bibliography*, e l'ISO, con gli standard sui thesauri e sul metodo di analisi, fissava punti imprescindibili riguardo al lessico e alle relazioni semantiche e riguardo all'analisi concettuale. D'altra parte le *subject headings* della Library of Congress, massima rappresentanza della tradizione, confermavano il proprio successo imponendosi, nonostante limiti riconosciuti, come modello sovranazionale anche in aree non anglofone, grazie all'attento lavoro di aggiornamento permanente e offrendo, non appena la telematica lo permise, la possibilità della catalogazione derivata, e sostituivano così varie esperienze autoctone, PRECIS incluso. Se in Italia la soggettazione in questi ultimi anni si è allontanata dal modello di riferimento considerato da Revelli, in una panoramica mondiale il suo lavoro resta spaventosamente attuale, a parte che per l'ovvia assenza dei dispositivi di memorizzazione e ricerca elettronica di cui oggi usufruiamo. E' un altro riconoscimento, che ha tutto il peso della supremazia mondiale, dove la nuova esperienza italiana resta minoritaria. Come se al curioso esploratore venisse restituita dalla comunità internazionale, per l'attenzione dedicatale, una sorta di conferma e approvazione, l'attestato di una sintonia nel porre sempre al centro il principio della praticità.

Un'ultima nota è per lo stile del testo. La quantità delle notizie fornite e delle argomentazioni addotte comporta una complessità del pensiero che è ben risolta nello stile tipico di Carlo Revelli, quello che conosciamo dai tanti suoi scritti e dalle esposizioni pubbliche che abbiamo ascoltato senza mai perdere il filo del discorso attraverso l'accumularsi armonico di subordinate e coordinate, di incisi, parentesi e amplificazioni, dove si articolano e sviluppano più linee tutte ben connesse come nella migliore tradizione polifonica. Un pregio raro, in tempi in cui abbondano le frasi brevi perentorie e ad effetto, le superficialità accostate o al contrario arzigogoli che si accartocciano su se stessi, un esempio di confronto critico e di chiarezza, di servizio alla comunicazione, in definitiva, a confermare un'attenzione sempre presente all'autore.

PINO BUIZZA, Biblioteca Queriniana, Brescia.
gbuizza@comune.brescia.it

Buizza, P. "Ri/leggere Revelli. Metonimia per una presentazione de *Il catalogo per soggetti* di Carlo Revelli". *JLIS.it* Vol. 3, n. 1 (Giugno/June 2012): 5616-1-5616-9. DOI: [10.4403/jlis.it-5616](https://doi.org/10.4403/jlis.it-5616). Web.

ABSTRACT: Il saggio ripercorre la storia e la struttura del volume "Il catalogo per soggetti" di Carlo Revelli, recentemente ripubblicato e diventato una pietra miliare nella storia della soggettazione in Italia. Il testo di Revelli, pur dovendo fare i conti con il cambiamento degli strumenti e dopo 40 anni dalla sua prima pubblicazione, rimane fondamentale per la comprensione e la messa in pratica della soggettazione nella catalogazione.

KEYWORDS: Catalogazione; Catalogo per soggetti; Indicizzazione per soggetto; Revelli, Carlo

Submission: 2012-05-06
Pubblicazione: 2012-06-01



JLIS.it



Dipartimento Scienze dell'antichità, Medioevo e Rinascimento e Linguistica

con il supporto di:

supported by:

Casalini
libri



Le Lettere

La piattaforma ICT, lo sviluppo e la manutenzione dell'installazione di OJS che ospita JLIS.it sono forniti da:

ICT platform, developing and maintenance for the OJS installation hosting JLIS.it are provided by:



Direttore Responsabile ai termini di legge: Nicola Cavalli
In attesa di iscrizione nel registro stampa del Tribunale di Milano.

Finito di stampare nel mese di luglio 2012 da

Ledizioni 
The Innovative LEDpublishing Company

<http://www.ledizioni.it>