



BIBLIOTECA DIGITAL

www.sciencia.cat

Preprints

The digital corpus of *Sciència.cat*: a work in progress*

Lluís Cifuentes i Comamala
Universitat de Barcelona
lluiscifuentes@ub.edu

ABSTRACT

The research strand regarding the dissemination of science and technology in Catalan during the Middle Ages and the Renaissance, initiated by the author in 1993 and continued at the Universitat de Barcelona from 2003 onwards, has embraced the exploitation of ICTs as a tool for the systematization, spreading and transfer of research. In 2006 the *Sciència.cat* website <<http://www.sciencia.cat>> was put online so that it might act as the showcase for that strand, and simultaneously as the framework within which the above-stated aims might be put into progressive effect, aims intended to take shape in a digital library and a relational database. The database was designed, constructed and financed in collaboration with the CODITECAM project, and data is being entered therein with a view to offering it to the public in early 2012. The contribution made thereby lays bare the origins of the project, as well as its characteristics, current situation and prospects for the future.

* This work falls within the framework of the research project 'Sciència.cat: hacia un corpus digital de la ciencia y de la técnica en catalán en su contexto románico durante la Edad Media y el Renacimiento' (FFI 2008-02163/FISO, 2009-2011) ['Sciència.cat: towards a digital corpus of science and technology produced in Catalan within the Romance context during the Middle Ages and the Renaissance], run by Lluís Cifuentes and funded by the Ministerio de Ciencia e Innovación (MICINN) <<http://www.sciencia.cat>>, as well as within that of the group 'Cultura i literatura a la baixa Edat Mitjana' (SGR 2009-1261), run by Lola Badia and recognized and funded by the Departament d'Innovació, Universitats i Empresa (DIUE) of the Generalitat de Catalunya <<http://www.narpan.net>>. The author is a member of the Centre de Documentació Ramon Llull (CDRL) as well as of the Institut de Recerca en Cultures Medievales (IRCV) at the Universitat de Barcelona. English translation by Robert D. Hughes.

Please, cite as: CIFUENTES I COMAMALA, Lluís, «The digital corpus of *Sciència.cat*: a work in progress», in Avenoza, Gemma – Soriano, Lourdes (eds.), *Humanities on the Web: The Medieval World = Humanitats a la xarxa: món medieval*, forthcoming (preprint). *Biblioteca Digital de Sciència.cat*, April 2012, Universitat de Barcelona (*date of accession*) <http://www.sciencia.cat/biblioteca/documents/Cifuentes_Hum.pdf>.

KEYWORDS

Sciència.cat – Digital corpus – Digital library – Database – Science – Technology – Middle Ages – Renaissance – Catalan

The digital corpus of *Sciència.cat*, which falls within a strand of research initiated by the author in 1993 and continued at the Universitat de Barcelona from 2003 onwards, aims to encourage study of the dissemination of science and technology in the Catalan language during the latter centuries of the Middle Ages and the early Renaissance within the framework of a process that has been called the vernacularization of knowledge. Such study has been made manifest through the elaboration of various monographs, among which stands out a general presentation of the process within the Catalan cultural sphere,¹ the editing of ancient texts as well as initiatives towards high popularization.² The analysis of this part of the Catalan cultural heritage has shown itself to be a valid and almost unique case study within the context of the European Latin West, and one that is helpful in understanding the aforementioned process.³

The organizing and cataloguing of the material which forms the object of study, namely, the scientific and technical works which circulated in the Catalan tongue (both as original compositions and as translations from other languages) between the thirteenth and sixteenth centuries, likewise constitutes a basic necessity for this strand of research. We have therefore chosen to exploit the possibilities offered by the new information and communication technologies (ICTs) as well as the so-called historical information systems (HISs), as tools for the systematization, spreading and transfer of research results.⁴ In 2006 the website *Sciència.cat* was put online so that it might act as the showcase for the research strand at the same time as the framework within which

¹ Lluís Cifuentes i Comamala, *La ciència en català a l'Edat Mitjana i el Renaixement*, 2nd ed. revised and expanded (Barcelona–Palma: Universitat de Barcelona–Universitat de les Illes Balears, 2006; originally published in 2002).

² Girolamo Manfredi, *Quesits o perquens (regiment de sanitat i tractat de fisiognomonía)*, ed. by Antònia Carré (Barcelona: Barcino, Els Nostres Clàssics B 25, 2004), is one of the critical editions which have been published. Jaume Roig, *Espill*, introductory study, [modernized] edition, translation into modern-day Catalan and commentary by Antònia Carré (Barcelona: Quaderns Crema, 2006), is an example of the previously mentioned high popularization. The same author has published the sixteenth-century Spanish translation of Manfredi's work and her critical edition of *Espill* within the Els Nostres Clàssics collection is currently in press.

³ See Lluís Cifuentes i Comamala, 'La vernacularización de la ciencia a finales de la Edad Media: un modelo explicativo a partir del caso del catalán', in *Aproximaciones al lenguaje de la ciencia*, ed. Bertha M^a. Gutiérrez Rodilla (Burgos: Fundación Instituto Castellano y Leonés de la Lengua, 2003), pp. 265–300 [Italian version: 'La volgarizzazione della scienza alla fine del medioevo: un modello interpretativo a partire dal caso del catalano', in *Filosofia in volgare nel Medioevo: atti del convegno della Società Italiana per lo studio del pensiero medievale (S.I.S.P.M.) (Lecce, 27–29 settembre 2003)*, ed. Nadia Bray and Loris Sturlese (Louvain-La-Neuve, Fédération Internationale des Instituts d'Études Médiévales, 2003), pp. 247–63].

⁴ See the account of the initial stages of this task of organizing and cataloguing as provided by Lluís Cifuentes i Comamala, 'Textes científiques en catalan (XIII^e–XVI^e siècles) dans les bibliothèques de France', *Médiévales: Langue, Textes, Histoire*, 52 (2007), 89–118; for the origins and characteristics of the digital corpus, see Idem, 'Sciència.cat: un corpus digital de la ciència i de la tècnica en català a l'Edat Mitjana i el Renaixement', in *La presència catalana a l'espai de trobada de la Mediterrània medieval: noves fonts, recerques i perspectives*, ed. Lluís Cifuentes, Roser Salicrú and Mercè Viladrich (Barcelona: IEMed, forthcoming).

such exploitation might be put into progressive effect, this taking shape in a digital corpus made up of a digital library and a relational database both of which will be described below.⁵

Our current staff consists of members of the funded research project whose strand of research receives finance therefrom, of a computer programmer, and a group of collaborators. Those forming part of the research project are: Lluís Cifuentes (Universitat de Barcelona), Antònia Carré (Universitat Oberta de Catalunya), Sebastià Giralt (Universitat Autònoma de Barcelona), Carmel Ferragud (Universitat Miguel Hernández d'Elx), Ilaria Zamuner (Università di Chieti) and Stefano Rapisarda (Università di Catania). The computer programmer is Miquel Hernández. Collaborators to date have included Miquel Raufast, Nin Sauleda, Roger Aluja, Teresa Huguet and Iván Armenteros.

I. THE DIGITAL LIBRARY

The digital library of *Sciència.cat* has been conceived with a view to facilitating systematized access to the primary and secondary sources which provide the focus of our research or which are of use to our particular research strand.⁶ It is structured in the following way:

- (i) *Catalogue of works*: a systematic, descriptive list, arranged according to spheres of knowledge, and detailing the scientific and technical works which circulated in Catalan between the thirteenth and the sixteenth centuries, both works which have been preserved and those which have been lost or whose location is currently unknown.⁷
- (ii) *Electronic editions*: a virtual library consisting of all the available works from among those listed in the 'Catalogue of works', these being arranged according to the same thematic principle. This section includes both editions and digital reproductions of the manuscripts or early printed books, and incorporates not only material prepared by our team of researchers but also – and now in particular – brings together systematically the existing resources on the internet.⁸
- (iii) *Bibliography*: a list of bibliographical references regarding science and technology within the Catalan cultural sphere and, in general, within the European and Mediterranean context which corresponded to it during the period in which this strand of research is interested, freely available on the internet and on the *Sciència.cat* server itself. This section, which currently (July 2011) includes more than 2,690 references, is provisional and aims to provide an indication of the bibliographical scope of the database that is presently under

⁵ *Sciència.cat: la ciència en la cultura catalana a l'Edat Mitjana i el Renaixement*, under the coordination of Lluís Cifuentes and Antònia Carré, 2006–, viewable at <<http://www.sciencia.cat>> in Catalan, Spanish and English versions.

⁶ Access available at: <<http://www.sciencia.cat/biblioteca/biblioteca.htm>>.

⁷ Published online at: <<http://www.sciencia.cat/biblioteca/catalegdobres.htm>>. This list brings the information available in Cifuentes, *La ciència en català* up to date.

⁸ Available at: <<http://www.sciencia.cat/biblioteca/edicions/00edelectron.htm>>.

construction (see below). Once the database is brought online, this section will become redundant.⁹

- (iv) *Sciència.cat Publications*: a space within which examples of research may be published which relate to Old Catalan scientific and technical texts and their cultural context – editions of works and pre-prints of articles in particular. This space is nurtured both by the researchers within the research strand themselves as well as by other researchers from outside who wish to present their work via a specialized resource. Works published within this space bear the seal of *Sciència.cat*.¹⁰
- (v) *Topics*: a section aimed at bringing about the high popularization of the topics characteristic of the research strand to which publication on the web assures the broadest dissemination, as well as at conveying to society the task being undertaken by our staff. Each of the selected topics is accompanied by the most up-to-date evaluation, as endorsed by the researchers responsible for such, as well as an indication of, and commentary upon, the highest quality internet resources available in that regard. This section, of interest to teachers, students and the educated public in general, is open to external collaboration.¹¹

We anticipate being able to increase the number of electronic editions available, of publications bearing our own seal and of the topics discussed with a broader audience in mind. One of the paths towards growth open to us is the digitalization of early editions and bibliography that are of particular interest in terms of the history of Old Catalan texts.

II. THE DATABASE

The other component of the digital corpus of *Sciència.cat* is a specialized database bearing the name 'Sciència.cat DB'.¹² This database is nourished, in part, by the material which makes up the digital library though, chiefly, by unpublished research on the primary sources. Its aim is to facilitate access to the content of such sources, and to do so in a systematic and relational fashion.

Sciència.cat DB has been built up in collaboration with the *Corpus digital de textos catalans medievals* (CODITECAM) funded research project, under the coordination of Professor Lola Badia at the Universitat de Barcelona (2006–2014), a project belonging to the same broad research group as does *Sciència.cat*.¹³ When it was first proposed, as an aspect of this project, to set up a network of relational databases aimed at classifying Old Catalan texts, a network whose focus would lie in works of an eminently literary nature (chansonniers and translations) or in authors closely connected

⁹ See: <<http://www.sciencia.cat/biblioteca/bibliografia/bibliografia.htm>>.

¹⁰ See: <<http://www.sciencia.cat/biblioteca/publicacionssc.htm>>.

¹¹ This section may be consulted at: <<http://www.sciencia.cat/biblioteca/temes/Otemes.htm>>.

¹² Presentation and future access via: <<http://www.sciencia.cat/db/bases.htm>>.

¹³ Information available at: <<http://www.narpan.net/recerca/projecte-coditecam.html>>. For the research group, see *Narpan: espai de literatura i cultura medieval*, under the coordination of Sadurní Martí and Miriam Cabré, 2003– <<http://www.narpan.net>>.

with literary history (Ramon Llull and Francesc Eiximenis),¹⁴ the writer of these pages was a staff member of this project. This fact enabled him to participate in the design and management of a technological platform of the utmost usefulness as regards the aims of the research strand being undertaken by *Sciència.cat*, which at that time was also making the leap to ICTs. Such collaboration between the CODITECAM and the *Sciència.cat* projects continues to this day as far as the design, development and financing of their common technological platform is concerned, and we share the same internal operational interface, programming and computer programmer.¹⁵

The *Sciència.cat* database aims to achieve the interrelated systematization of information within the seven fields described below, in whose regard it marries both the current situation and the prospects for the future:

- (i) *Works*: provides a catalogue of the scientific and technical works written in – or translated into – Catalan during the thirteenth to the sixteenth centuries. It systematizes the general information relating thereto in terms of the latter's identification (e.g. authors, titles, translators, dates), content (e.g. incipit, explicit, complementary material), transmission (i.e. manuscripts, printed editions, documentary evidence), dissemination (e.g. interpolations) and bibliography (i.e. editions, secondary bibliography, digital texts). Data wrapping has been completed and, once the material has been checked, we expect the catalogue to be available for online consultation at early 2012.
- (ii) *Books*: provides a catalogue of the book-related sources (i.e. manuscripts and early printed books) which contain the works. It systematizes the codicological information pertaining thereto on the basis of a direct analysis of the volumes, and gives an external description (i.e. the volume's identification, binding, history, bibliography, dating and origin, material issues, page layout and script) as well as an internal one (i.e. details of the works it contains), as is customary in such catalogues. Work is currently being carried out on data wrapping within this field, a lesser part of which will be visible at early 2012, after the appropriate checking has been done, the whole of which awaits subsequent completion.
- (iii) *Documents*: will be a catalogue of documents concerning the works, their authors, translators or other individuals involved and their context. It will include chiefly (though not exclusively) documentation regarding books and their readers – the documentation itself coming above all from the inventories of

¹⁴ A presentation of the chansonnier database is given by Miriam Cabré and Sadurní Martí, 'Per a una base de dades dels cançoners catalans medievals: l'exemple de *Sg*', in *Actes del Tretzè Col·loqui Internacional de Llengua i Literatura Catalanes (Universitat de Girona, 8–13 de setembre de 2003)*, ed. Sadurní Martí *et al.* (Barcelona: Associació Internacional de Llengua i Literatura Catalanes–Publicacions de l'Abadia de Montserrat–Universitat de Girona: Institut de Llengua i Cultura Catalanes, 2007), III, pp. 171–86. The database of literary translations (*Translat*) is discussed in Josep Pujol, 'Translat: una base de dades de traduccions literàries al català medieval (1300–1500)', *Quaderns: Revista de Traducció*, 15 (2008), 213–21. The *Base de dades Ramon Llull–Llull DB*, dir. by Anthony Bonner and coord. by Lola Badia and Albert Soler, 2002–, which may be consulted at: <<http://orbita.bib.ub.edu/llull/>>, will ultimately retain its current format. A funded project has recently been approved which, during the three-year period between 2012–2014, will add another database to this network, namely one devoted to Arnau de Vilanova (Arnau DB), a project which will be undertaken by a team under the direction of Sebastià Giralt (Universitat Autònoma de Barcelona).

¹⁵ Thanks to this common technological platform, it is expected that certain searches may be carried out conjointly, a feature which is of particular interest in the case of *Translat* and of *Sciència.cat*.

- old libraries – a fact which enables one to extend and complete the list of works which have been preserved by means of information concerning those which circulated but have not survived. At present, a provisional model for catalogue entries has been designed which we expect to be able to implement from 2012 onwards.
- (iv) *Names*: is a catalogue containing the personal names and the place names present in the works and in the primary sources which form the object of study. It systematizes the biographical data relating to the individuals involved in the composition, translation, copying or possession of the works (e.g. authors, translators, copyists, correctors, editors, and those who owned the works) as well as place names. It takes account of formal and linguistic variants and is able to include external links. Work is being conducted on this field in parallel with the other fields which are already in place or are in the process of data wrapping.
 - (v) *Bibliography*: is a catalogue of the editions of and studies concerning the works or their context, a characteristic which turns it into a broad bibliography of the history of science and technology in the Medieval West and the early Renaissance, especially in the Mediterranean basin. At present, almost 10,000 entries have been introduced, some of which (more than 2,690 entries) are linked to the full text as freely available on the internet or on our own server. As has previously been stated, such cases are already accessible within the ‘Digital Library’. This field will be made public in early 2012.
 - (vi) *Vocabulary*: will be a catalogue of the technical vocabulary present in the works which form the object of study, the aim of which is to constitute a specialized and up-to-date complement to other more general initiatives relating to the Catalan language in its ancient forms.¹⁶ This undertaking will start out by covering the medical sphere, although we anticipate that it may progressively be extended to include other fields, depending on the availability of texts and the researchers who collaborate thereupon, since we intend it to become a realm that is open to external collaborators. At present, the model we have designed for catalogue entries is being implemented and we expect the work of data wrapping to begin during the three years between 2012–2014.¹⁷
 - (vii) *Images*: will be a catalogue of illustrations concerning science and technology in the Middle Ages and the early Renaissance, for purely educational and non-commercial purposes. It seeks to be a support tool for teaching, for academic gatherings and for publications. Even though the *Sciència.cat* team has been putting together material pertinent to this field, there are currently no plans to develop this area.

¹⁶ In particular, the *Diccionari català-valencià-balear (DCVB)*, a work initiated by Antoni M. Alcover, and edited by Antoni M. Alcover and Francesc de B. Moll, with the collaboration of Manuel Sanchis Guarnier and Aina Moll (Palma: Moll, 10 vols.; 2nd ed. corrected and updated: 1964-1969), accessible at: <<http://dcvb.iec.cat>>; and the *Vocabulari de la llengua catalana medieval*, by Lluís Faraudo de Saint-Germain, electronic edition under the direction of Germà Colón Domènech (Barcelona: Institut d’Estudis Catalans, 2007–), accessible at: <<http://www.iec.cat/faraudo>>.

¹⁷ A specific presentation of this is given by Lluís Cifuentes i Comamala, ‘Hacia un nuevo Vocabulario del catalán técnico medieval’, in *Lengua de la ciencia y diccionarios: actas de las III Jornadas de la red temática «Lengua y ciencia» (Coimbra, 28–30 octubre 2009)*, ed. Graça Rio-Torto (Coimbra: Universidade de Coimbra, in press).

As can be seen, our efforts in general are directed towards a staggered availability to the public from early 2012 onwards. Within the framework of the technological and financial collaboration established with the aforementioned CODITECAM project, the user interfaces have already been designed, as have those appropriate to each database within this project and the interface particular to *Sciència.cat*, while the first databases themselves, though presently in a more or less embryonic state, have already been published on the internet.¹⁸

III. CONCLUSIONS

The digital corpus of *Sciència.cat* has the following principal characteristics: it is made up, for the most part, of unpublished material resulting from pure research; it offers information that is suitably classified and powerfully interrelated thanks to the fact that it exploits new technologies; it is openly accessible to the academic community and to the educated public in general (when the quantity and quality of the data that have been wrapped allows it). Characteristics such as these will mean that this digital corpus can function as an operational tool of the first order, chiefly for the researchers responsible for the corpus themselves though also for the rest of the academic community. In this way it will guarantee an extraordinary revitalization and dissemination of the Catalan historico-cultural legacy, the content of which and the analyses derived therefrom will form, with greater ease, an integral part of international research. From its inception, it aimed to become an obligatory point of reference for scholars of the disciplines involved, and for the moment the index of visitors to the *Sciència.cat* website would seem to corroborate this.

For all these reasons, this corpus constitutes one of the most important initiatives as regards the utilization and dissemination of the Catalan cultural legacy, and offers evident and immense opportunities for research, teaching and popularization.

IV. BIBLIOGRAPHY

Base de dades Ramon Llull/Llull DB, dir. by Anthony Bonner and coord. by Lola Badia and Albert Soler, 2002– <<http://orbita.bib.ub.edu/llull/>>.

Miriam Cabré and Sadurní Martí, ‘Per a una base de dades dels cançoners catalans medievals: l’exemple de *Sg*’, in *Actes del Tretzè Col·loqui Internacional de Llengua i Literatura Catalanes (Universitat de Girona, 8–13 de setembre de 2003)*, ed. Sadurní Martí *et al.* (Barcelona: Associació Internacional de Llengua i Literatura Catalanes–Publicacions de l’Abadia de Montserrat–Universitat de Girona: Institut de Llengua i Cultura Catalanes, 2007), III, pp. 171–86.

¹⁸ See *Cançoners DB*, coord. by Miriam Cabré and Sadurní Martí, 2011– <<http://www.candb.narpan.net/>>; *Translat DB*, coord. by Lluís Cabré and Josep Pujol, 2011– <<http://www.translatdb.narpan.net/>>; and *Eiximenis DB*, coord. by Xavier Renedo and Sadurní Martí, 2011– <<http://www.eiximenis.narpan.net/>>.

- Cançoners DB*, coord. by Miriam Cabré and Sadurní Martí, 2011–
<<http://www.candb.narpan.net/>>.
- Lluís Cifuentes i Comamala, ‘La vernacularización de la ciencia a finales de la Edad Media: un modelo explicativo a partir del caso del catalán’, in *Aproximaciones al lenguaje de la ciencia*, ed. Bertha M^a. Gutiérrez Rodilla (Burgos: Fundación Instituto Castellano y Leonés de la Lengua, 2003), pp. 265–300.
- Lluís Cifuentes i Comamala, ‘La volgarizzazione della scienza alla fine del medioevo: un modello interpretativo a partire dal caso del catalano’, in *Filosofia in volgare nel Medioevo: atti del convegno della Società Italiana per lo studio del pensiero medievale (S.I.S.P.M.) (Lecce, 27–29 settembre 2003)*, ed. Nadia Bray and Loris Sturlese (Louvain-La-Neuve, Fédération Internationale des Instituts d’Études Médiévales, 2003), pp. 247–63.
- Lluís Cifuentes i Comamala, *La ciència en català a l’Edat Mitjana i el Renaixement*, 2nd ed. revised and expanded (Barcelona–Palma: Universitat de Barcelona–Universitat de les Illes Balears, 2006; originally published in 2002).
- Lluís Cifuentes i Comamala, ‘Textes científiques en catalan (XIII^e–XVI^e siècles) dans les bibliothèques de France’, *Médiévales: Langue, Textes, Histoire*, 52 (2007), 89–118.
- Lluís Cifuentes i Comamala, ‘*Sciència.cat*: un corpus digital de la ciència i de la tècnica en català a l’Edat Mitjana i el Renaixement’, in *La presència catalana a l’espai de trobada de la Mediterrània medieval: noves fonts, recerques i perspectives*, ed. Lluís Cifuentes, Roser Salicrú and Mercè Viladrich (Barcelona: IEMed, forthcoming).
- Lluís Cifuentes i Comamala, ‘Hacia un nuevo Vocabulario del catalán técnico medieval’, in *Lengua de la ciencia y diccionarios: actas de las III Jornadas de la red temática «Lengua y ciencia» (Coimbra, 28–30 octubre 2009)*, ed. Graça Rio-Torto (Coimbra: Universidade de Coimbra, in press).
- Diccionari català-valencià-balear (DCVB)*, initiated by Antoni M. Alcover, and edited by Antoni M. Alcover and Francesc de B. Moll, with the collaboration of Manuel Sanchis Guarner and Aina Moll, 2nd ed. corrected and updated (Palma: Moll, 10 vols., 1964–1969) <<http://dcvb.iec.cat/>>.
- Eiximenis DB*, coord. by Xavier Renedo and Sadurní Martí, 2011–
<<http://www.eiximenis.narpan.net/>>.
- Lluís Faraudo de Saint-Germain, *Vocabulari de la llengua catalana medieval*, electronic edition under the direction of Germà Colón Domènech (Barcelona: Institut d’Estudis Catalans, 2007–) <<http://www.iec.cat/faraudo>>.
- Girolamo Manfredi, *Quesits o perquens (regiment de sanitat i tractat de fisiognomonía)*, ed. Antònia Carré (Barcelona: Barcino, Els Nostres Clàssics B 25, 2004).
- Narpan: espai de literatura i cultura medieval*, coord. by Sadurní Martí and Miriam Cabré, 2003– <<http://www.narpan.net/>>.
- Josep Pujol, ‘Translat: una base de dades de traduccions literàries al català medieval (1300–1500)’, *Quaderns: Revista de Traducció*, 15 (2008), 213–21.
- Jaume Roig, *Espill*, introductory study, [modernized] edition, translation into modern-day Catalan and commentary by Antònia Carré (Barcelona: Quaderns Crema, 2006).
- Sciència.cat: la ciència en la cultura catalana a l’Edat Mitjana i el Renaixement*, coord. by Lluís Cifuentes and Antònia Carré, 2006– <<http://www.sciencia.cat/>>.
- Translat DB*, coord. by Lluís Cabré and Josep Pujol, 2011– <<http://www.translatdb.narpan.net/>>.