The Challenges of Developing a Common Glossary Across Different Countries and Languages in InterPARES 3: Some Examples from the TEAM Brazil Experience

Claudia Lacombe Rocha; National Archives of Brazil; Rio de Janeiro, Brazil[1]

Abstract

The InterPARES 3 Project (International Research on Permanent Authentic Records in Electronic Systems) is an international collaborative research endeavor composed of numerous regional, national, multinational and multidisciplinary research teams. In such a varied context, the precision and consistency of the terminology used in the course of the research is vital to the success of the Project. Accordingly, the Project is developing a multilingual Terminology Database, the primary purpose of which is to facilitate communication and research among researchers in different administrative and disciplinary contexts and across cultural boundaries by defining terms and relationships among terms used in the research. The Terminology Database is composed of two main components: the Glossary and the Dictionary. The Glossary contains the definitive definition for the key terms as they are used in the Project's working and published documents. As such, the Glossary terms with their definitions are the key to the communication of the findings of the Project. Because many of the terms in the Glossary have multiple cross-disciplinary, administrative and/or cross-cultural definitions, it is important to account for plurality of meaning, which is precisely what the Dictionary is intended to do. Thus, the Dictionary contains other definitions of Glossary terms in cases where those terms are used differently and/or have different definitions in other contexts.

InterPARES 3 researchers are facing many challenges with the Glossary translations to the languages of each research team. The difficulties and pitfalls are due to many reasons: imprecision of definitions and national uses of them, differences in the governmental and juridical systems of each country and their different administrative practices, nonexistence of certain archival practices in some countries, and so on. Two rounds of translations have already been completed, making a total of 112 translated terms. It is anticipated that the final Glossary will contain more than 400 terms. This paper examines some of the key issues and challenges surrounding the Project's Glossary translation effort, illustrated by some cases dealt with by TEAM Brazil researchers.

Introduction

The InterPARES Project (International Research on Permanent Authentic Records in Electronic Systems) is an international, collaborative research effort aimed at the development of knowledge essential to the long-term preservation of authentic records created and/or maintained in digital form. The Project involves researchers from many countries and disciplines. In such a varied context, the precision and consistency of the terminology used in the course of the research is vital to the success of the Project. Accordingly, InterPARES 2 researchers developed a terminology database composed of two main instruments: a Glossary and a Dictionary.

The Glossary is a list of terms and definitions that is crucial to an understanding of records creation, keeping and preservation, and presents the definitive definitions for the terms as they are used in the Project's working and published documents. As such, the Glossary terms with their definitions are the key to the communication of the Project findings.

Because some terms can have multiple definitions corresponding to different disciplines or cultural and administrative contexts, it is important to register this plurality of meaning in another instrument. Thus, the Dictionary contains other definitions of Glossary terms in cases where those terms are used differently from the standpoint of Archival Science, Computer Science, Library and Information Science, Arts, etc.

The InterPARES 3 Project is developing a multilingual Terminology Database. Researchers are translating the <u>terms</u> and their <u>definitions</u> in the InterPARES Glossary to each TEAM language. The primary purpose of this task is to facilitate communication among researchers by taking into consideration cultural boundaries and differences. That is a very important task, since, to disseminate InterPARES theory, methodology and findings, we must use a "common language" and a precise meaning to basic concepts, especially when involving different countries and different languages.

Translating the InterPARES Glossary has being a big challenge. In many cases a term cannot be translated literally because the word may have a different meaning than the one used in the archival practice of a specific country; in some other cases there is no corresponding term used in the country because of the differences in the archival practices. It is not sufficient to simply translate the term; it must be defined and explained in the country's technical language.

---

[1] This paper registers the work of the TEAM Brazil Terminology Task Force, which is comprised of the following individuals: Brenda Rocco, Carlos Ditadi, Claudia Lacombe Rocha, José Márcio Rangel, Margareth da Silva, Rosely Rondinelli and Vanderlei dos Santos.

Many terms are only understandable in another language after thorough analysis of the professional, cultural, legal, historical and administrative context of the term. A dictionary is not enough for translating a term, the understanding of a term needs the understanding of the conceptual framework where it was adopted.

In 1984, when the International Council on Archives presented the first version of the *Dictionary of Archival Terminology (DAT)*, with definitions of archival terms in English and French, plus translation of terms to five other languages (German, Spanish, Italian, Dutch and Russian), Michel Duchein had already highlighted some difficulties and pitfalls that are commonly faced when doing this type of translation effort. In his article "Archives into Tower of Babel: international terminology problems," Duchein said that "These difficulties come from three sources: first the frequent imprecision of definitions and national uses of them; secondly the increasing vocabulary divergence within the same language in homophone countries; and finally, the fact that archival science is closely connected to the governmental and administrative juridical system of each country and therefore the vocabulary reflects a set of concepts that is difficult to transpose from one country to another."[2]

When translating the InterPARES Glossary, TEAM Brazil researchers faced these difficulties and pitfalls noted by Duchein, and also some others. Two rounds of terms have already been translated by all TEAMs, making a total of 112 terms. Some cases can be presented to illustrate the problems encountered in this effort.

Record: a classic pitfall

In the InterPARES 2 Dictionary, there are fourteen different definitions for the term **record** as used in various disciplines and some of them are quite different from the one adopted by the Project, which defined **record** as "A document made or received in the course of a practical activity as an instrument or a byproduct of such activity, and set aside for action or reference."[3]

A **record** is not any document, but only those created by a physical or juridical person that <u>participates in an action</u> (that is, is created to support the action or to register the action and is therefore a natural byproduct of the action) and that is <u>set aside</u>. Translating the term **record** to Latin languages was not easy at first. Traditionally, in Brazil, the translation adopted for **record** by archivists is "*Documento arquivístico*" or "*Documento de arquivo*," which qualifies the document as an archival document or a document that pertains to an archives. A similar term is used in French (*Document d' archives)*,[4] Spanish (*Documento de archivo*) and Italian (*Documento archivistico*).[5]

In fact, the translation of **record** as "*Documento arquivístico*" was recently adopted in Brazil, and it is important to note that InterPARES Project findings strongly contributed to that change. Previously, it was more common to refer to these entities simply as "*documents*," making no distinction between **documents** and **records**. In Brazil, there are two important dictionaries published by archival institutions: the *Dictionary of Archival Terminology* published by the Association of Brazilian Archivists in 1996[6] and the *Brazilian Dictionary of Archival Terminology* published by the National Archives in 2005.[7] In the first one, the term "*Documento arquivístico*" is presented as equivalent to the English term **archives**, while the second one simply ignores the term. Only when the archival description standards of the International Council on Archives were translated in Brazil—ISAD(G) in 2001 and ISAAR(CPF) in 2004—was the term "*Documento arquivístico*" presented in the glossary as equivalent to **record**.[8]

It is also worth noting that the term "*Documento arquivístico*" is used in Brazil for both **record** and **archives**, while Americans often use the term **record** only for current and semi-current documents and reserve the term **archives** or **archival documents** for non-current documents that are preserved due to their enduring or secondary value. This divergence reflects the differences of archival practice in these countries. In the United States, two different professions were constituted: records management, which takes care of records in the creator's environment, and archives management, which is responsible for managing and preserving the archives (documents with historical or permanent value). Each one is organized independently and has specific certification processes and independent professional associations. Differently, in Brazil and many other countries, archivists are professionals who take care of documents throughout their lifecycle, from their creation in organizations to their custody with preservers.

For archivists, the understanding and translation of the terms **records** and **archives** are no more questionable but still pose some problems when involving other professionals.

---

[2] Michel Duchein (2007), "Os arquivos na torre de Babel*," Acervo* 20: 15. Originally published in 1985 in "Les archives dans la Tour de Babel: problèmes de terminologie internationale," *Gazette des Archives* 129: 103-113. The citation was translated to English by the author of this paper.

[3] InterPARES 2 Terminology Database. http://www.interpares.org/ip2/ip2_terminology_db.cfm.

[4] International Council on Archives, *DAT III* [Dictionary of Archival Terminology]: english list (draft). Complied by the Group on Terminology of the ICA. http://www.staff.uni-marburg.de/~mennehar/datiii/engterm.html.

[5] InterPARES 3 Glossary. http://www.interpares.org/rws/rws_terminology.cfm?status=wglossary. (restricted access to InterPARES researchers)

[6] Ana Maria de Almeida Camargo and Heloísa Liberalli Bellotto, comp., *Dicionário de terminologia arquivística* (São Paulo: Associação dos Arquivistas Brasileiros, São Paulo: Secretaria de Estado da Cultura, 1996).

[7] Arquivo Nacional, *Dicionário Brasileiro de Terminologia Arquivística* (Rio de Janeiro: Arquivo Nacional, 2005).

[8] Conselho Internacional de Arquivos, *ISAD(G): Norma geral internacional de descrição arquivística*, trans. Arquivo Nacional (Rio de Janeiro: Arquivo Nacional, 2001), 4; Conselho Internacional de Arquivos, *ISAAR(CPF): Norma internacional de registro de autoridade arquivística para entidades coletivas, pessoas e famílias*, trans. Arquivo Nacional (Rio de Janeiro: Arquivo Nacional, 2001), 14. These archival description standards from the International Council on Archives (ICA) were translated by a task force at the National Archives in Brazil; it is not an official version of the ICA document but it is an important reference for Brazilian archivists.

The translation of the word **record**, as commonly used in Portuguese, would be *registro* or *gravação*, and indeed we must note that *registro* or *gravação* are related to other definitions of the term **record** present in the InterPARES 2 Dictionary. They have different meanings from the one used in Archival Science and are related to other fields or disciplines and therefore were not adopted in the InterPARES Glossary. Some examples are:

> *n.*, In data processing, a grouping of interrelated data elements forming the basic unit of a file.

> *n.*, Any electronic, photographic or mechanical recording of music, singing, dialogue, sound effects or visual events, including CDs, DVDs, audio tapes, films, videos and the like.

> *n.*, An ordered set of fields, usually stored contiguously. The term is used with similar meaning in several different contexts. In a file, a "record" probably has some fixed length, in contrast to a "line" which may have any length and is terminated by some End Of Line sequence. A database record is also called a "row". In a spreadsheet it is always called a "row". Some programming languages use the term to mean a type composed of fields of several other types (C calls this a "struct").

For all these meanings of **record**, the corresponding term in Portuguese would actually be *registro* or *gravação*. This fact causes considerable confusion when people from other fields are translating an article or a presentation in an event about archives. Instead of translating **record** as *documento arquivístico*, it is commonly given as *registro*. This happens very frequently when people from other fields related to archives, such as Information Technology professionals, are speaking about "Electronic Recordkeeping" or "Digital Preservation."

A similar problem occurs with the use of the term **record** in English; although there is not a translation issue, professionals from other fields than archives understand the term differently. As suggested by the many definitions for the term that appear in the InterPARES 2 Dictionary, there are many different understandings of the term **record** in different disciplines. Eric Ketelaar highlights this matter when he questions "How can we as a profession communicate with, for example, IT people, librarians and museum curators, who understand the term **record** differently?"[9]

Digital records need the involvement of many actors from different fields. Although the right translation is well known by archivists, it is not for other professionals. It is vital to disseminate the right translation of **record** in the archival field (and even the understanding of the term **record** in English) as this is the first and basilar concept to our work. We are dealing with *Documentos arquivísticos* and not only *Registros*.

Records Management and Recordkeeping

The difficulties faced in the translation of the term **recordkeeping** are related to the differences in traditions and practices in different countries, which consequently presents the problem that a term used in one country may not have a corresponding one in another country.

As defined in the InterPARES 3 Glossary, **recordkeeping** is "the function of capturing, storing and maintaining records and information about them, and the set of rules governing such function." And also, in the InterPARES 2 Chain of Preservation model, the function of recordkeeping is well represented as part of the records management function, which includes the making and the keeping of records.[10] Once the record is created it is transferred from the record-making system to the recordkeeping system, the latter of which is responsible for maintaining the authenticity of the record.

The obstacle presented is to find a term in Portuguese to translate **recordkeeping** and the heart of this problem is born in the differences in records management practices in the various countries.

According to Joaquim Llansó i Sanjuan, the United States of America was the first country to elaborate the concept of a Record Management System, since the 1940s, followed by the emergence of many other national "models" of record management in other countries, related to each context. The American approach for records management was more influenced by an administrative and economic motivation, while in Europe the problem was treated specially from the standpoint of the historical interest of documents. The American model was primarily a method to optimize the performance of administration and to limit the quantity of records created and the period of time that they were meant to be kept.[11]

Nevertheless, despite the differences of concept and practices performed in many countries in relation to American Records Management, the development of a systematic administration of records between their creation and disposition was known by the generic designation of **records management**. Because of these differences, in the beginning, archivists had some difficulties in translating **records management**, but now there is no discussion about the adoption of "*Gestão de documentos*" in

---

[9] Eric Katelaar (1997), "The Difference Best Postponed? Cultures and Comparative Archival Science," *Archivaria* 44: 142.

[10] See Terry Eastwood, Hans Hofman and Randy Preston, "Part Five—Modeling Digital Records Creation, Maintenance and Preservation: Modeling Cross-domain Task Force Report," [electronic version] in *International Research on Permanent Authentic Records in Electronic Systems (InterPARES) 2: Experiential, Interactive and Dynamic Records*, Luciana Duranti and Randy Preston, eds. (Padova, Italy: Associazione Nazionale Archivistica Italiana, 2008), 2-46. http://www.interpares.org/ip2/display_file.cfm?doc=ip2_book_part_5_modeling_task_force.pdf.

[11] Joaquim Llansó i Sanjuan, *Gestión de Documentos. Definición y Análisis de Modelos* (Bergara: IRARGI, Centro de Patrimonio Documental de Euskadi, Departamento de cultura, Gobierno Vasco, 1993).

Portuguese, "*Gestion de documents*" in French and "*Gestión de documentos*" in Spanish, related to this generic definition of **records management**.

In Brazil, the concept of records management emerged in the 1960s after the translation of Theodore Schellenberg's book, *Modern Archives*.[12] The records management practices developed by Marilena Leite Paes at "Fundação Getúlio Vargas," a private institution, was the first important initiative in this field assumed by an organization in Brazil, and the publication of Marilena Leite Paes' book, *Arquivo: Teoria e Prática*,[13] was also an important milestone in records management in Brazil.

In 1975, a Pre-archive Office was created at the National Archives of Brazil. It was the first initiative related to taking care of semi-current records, but the activities developed in this office were not yet oriented by a records management theory or methodology. It was not until the 1980s that an office was created at the National Archives of Brazil to implement a records management program in the government administration. In 1990s, the National Archives and the National Council on Archives together developed guidelines and instruments to guide all records management procedures in Government Administration. Some examples are the classification code, records schedules[14] and guidelines for transferring and destroying records. Legally, the National Archives is responsible for records management in all of the Federal Government, and the guidelines developed determine that the disposition of all records created by government offices must be authorized by the National Archives and publicized before being destroyed.

These official instruments, the legislation and also the main Brazilian archival dictionaries define **records management** as "all the procedures and technical operations related to the production, flow, use, appraisal and archiving of current and semi-current records aiming at their destruction or transfer for an Archives." This is largely in agreement with the InterPARES definition of **records management**. The only difference resides in the idea, within InterPARES, of two sub-systems (record-making and recordkeeping) under the umbrella of a records management system. Brazilian practice in records management does not split it into two sub-systems. Rather, records management is seen as a <u>continuous flow</u> from creation to disposition. From this standpoint there is one single system that manages the records throughout all the phases of records management: creation, use and maintenance and disposition.

The idea of distinct record-making and recordkeeping systems was not developed in Brazil and consequently there are no corresponding words or terms for these concepts used by Brazilian archivists. Consequently, new terms need to be created in Portuguese to address the understanding of records management and its sub-systems as used in InterPARES.

In Portuguese, the word **keeping** is *manutenção*, which is similar to **maintain** in English. From one standpoint it seems to be a good choice as it is in agreement with the translation suggested in other Latin languages (e.g., *mantenimiento* in Spanish or *tenuta* in Italian). But from another standpoint this option seems less ideal since the function of "records maintenance" is only one of the functions involved in **recordkeeping**, which embodies other functions like: storage, use and disposition. A better choice would be a term that embraces all the functions or at least that is not restricted to one singular function of recordkeeping.

Even the creation of a term involves some difficult choices and requires a lot of research and study. In some cases, it may be necessary to promote greater debate outside InterPARES boundaries; i.e., involving the country's archival community. Actually, the differences of practices may create obstacles to the task of translating. Ketelaar observes one such obstacle created by differences of practices and cultures, noting that "When translating a word, an expression, a term into a foreign language, we try to convey the deeper cultural meaning, but the equivalent can only be an exact rendering of the original if the two cultures are congruent." "But," he further muses, "are there any congruent cultures?"[15]

Create and make records

Another kind of obstacle can be posed by the inconsistent use of a word. The translation of **create** and all its derivations (creator, create record, record creation) presented a problem when compared to the translation of the term **make**. Both are referenced in InterPARES documents with a very specific meaning and it is important that they are properly translated as they are used in many documents produced within the InterPARES context.

The definitions of **creator** and related terms as they appear in the InterPARES 3 Glossary are:

Term	InterPARES 3 Glossary Definition
creator	The physical or juridical person who makes, receives, or accumulates records by reason of its mandate/mission, functions or activities.
create record	To make and set aside or receive and set aside records.
record creation	The first phase of a record's lifecycle in which a record is made or received and then set aside for action or reference.

[12] Theodore R. Schellenberg, *Arquivos Modernos. Princípios e Técnicas*, 2nd ed. (Rio de Janeiro: Editora FGV, 2002).

[13] Marilena Leite Paes, *Arquivo: Teoria e Prática*, 3rd ed. (Rio de Janeiro: Editora FGV,1997).

[14] In 2001, the National Council on Archives published a Classification Code and a Record Schedule for records related to administration activities, common to all government offices. These instruments were adopted by the National Archives, and all Federal Government Administrations must use them. The publication also presents the design to guide government offices for developing a classification code and record schedule for the records related to their specific activities.

[15] Katelaar, "The Difference Best Postponed," ibid., 144.

All of these definitions present creation as a set of actions that includes making, receiving and setting aside records. The usual literal translation to Portuguese would be the words *criador / criar / criação*, but these are not used in Brazil for documents nor for records; instead, the terms used by archivists are *produtor / produzir / produção*. This translation is adopted in Brazilian standards and dictionaries of archival terms.

In the Brazilian translations of ISAD(G) and ISAAR(CPF), the term **creator** was translated as *produtor*, meaning "the person or institution that made, accumulated and/or kept the records during the management of her/its activities."[16] That is in accordance with the InterPARES definition.

In the Brazilian dictionary of archival terminology, there is the term "*Entidade produtora*," meaning "an entity or organization, a person or a family identified as generating an archive. Also called producer."[17]

The main problem is that Brazilian archivists also use the word *produzir* to mean "make records." This use is presented in legislation and archives manuals where, for the expression "records made or received," it is always said "*documentos produzidos ou recebidos*," and, for the expression "making of records," it is also said "*produção de documentos*."

In the InterPARES 2 Glossary, the term **record-making** is defined as "The whole of the principles, policies, rules and strategies that controls the process of creating records from made or received documents." Creating a record includes the process of making a record or the process of receiving a record and setting aside the record. In the InterPARES context, in relation to records, the sense of the term **create** is broader than the term **make**.

The meaning of the term *produzir* as used in Brazil changes depending on the context in which it is employed by archivists: it can be **create** (make or receive and set aside) or simply **make**. This inconsistent use of the term *produzir* by Brazilian archivists never seemed to be a problem since, usually, the different meanings could be identified depending on the context in which the term was used. However, the task of the InterPARES 3 Glossary translation highlighted this inconsistency since the terms **create** and **make**, as adopted in the Project, are distinct.

At this point Brazilian researchers must make a choice:

a) Either translate both terms, **create** and **make**, using the same word in Portuguese (*produzir*). This way the traditional use of the terms in Brazil is maintained, or

b) Try to correct the traditional usage of the word "*produzir*" by proposing, for example, another translation for **make** record, such as "*gerar documento*" or "*elaborar document*."

Paraphrasing Eric Katelaar, "Should we postpone and conceal the inconsistency of the use of "*produzir*" for different meanings?"[18]

The use of terms in English – no translation

Another issue faced by researchers was the routine use of terms in English or, more often, the acronym of the English term by professionals in Brazil, typically when related to IT terms. It is very common in Brazil for IT professionals to use technical terms in English even when writing articles or technical materials.

Although it may seem a little peculiar, in these cases there is no choice but to simply adopt the English term or the acronym in the Brazilian version of the InterPARES 3 Glossary. Nevertheless, it must be noted that this is not a general practice; the majority of IT terms in the Glossary are in fact translated to Portuguese.

Some examples to illustrate this case are:

Term	InterPARES 3 Glossary Definition	Term as used in Brazil
data grid	*n.*, The registration of digital entities into a logical name space. Manipulations of registered material can then be automated through any standard computer application programming interface (API).	data grid
document type definition	*n.*, (DTD) The definition of a document type in SGML or XML, consisting of a set of mark-up tags and their interpretation.	DTD
digital videodisc	*n.*, (DVD) A type of single- or double-sided, optical digital medium that is capable of storing from 4.7 to 8.5 gigabytes of digital data on two continuous, microscopic, spiral tracks or grooves that are cut and read by a laser beam. Its logical format specifications are governed by the Universal Disk Format (UDF) specification.	DVD
digital audio tape	*n.*, (DAT) A type of magnetic digital medium that can store up to 4 gigabytes of digital data per cassette by using helical scan recording.	Fita DAT

---

[16] Conselho Internacional de Arquivos, *ISAD(G),* ibid., 5; Conselho Internacional de Arquivos, *ISAAR(CPF)*, ibid., 14.

[17] Arquivo Nacional, *Dicionário Brasileiro*, ibid., 84.

[18] Katelaar, "The Difference Best Postponed," ibid., 142. The author poses a similar question about the differences on the use of the terms "records" and "archives" by Americans and Englishmen: "Should we postpone and conceal the differences in our discourse on the nature and administrations of records and archives?"

Conclusion

The cases presented in this paper show how complex is the task of translating a glossary. The importance of the decisions made in the course of translating the InterPARES 3 Glossary is enormous, because all material produced by researchers in the InterPARES context should then be translated to the various languages of each TEAM according to the InterPARES 3 Multilingual Glossary.

Moreover, as reported in the cases noted above, some issues call attention to inconsistencies, differences of some uses and practices, and the absence of other practices and concepts. Stimulated by this, researchers needed to go deeper in study and to start a debate.

Duchein, in the article cited at the beginning of this paper, remarks that "the study of terminological problems is one of the keys for archival science. It helps to get notions more precise, to disperse empiricism vapor and obliges us to question some certitudes acquired by routine."[19]

The terminology task in InterPARES 3 is not only developing a common language for InterPARES researchers, it also is helping the researchers from each regional, national and multinational TEAM to revise the archival concepts and terms used in their own country.

Author Biography

Claudia Lacombe Rocha received a bachelor degree in History from Pontifícia Universidade Católica do Rio de Janeiro (1987), and an MA in Informatics from Universidade Federal do Rio de Janeiro (2003). In 1989 she began to work at the National Archives in Brazil and since 2003 she has been in charge of the Electronic Records Management and Preservation Task Force at the National Archives. She is also President of the Electronic Records Committee at the National Council on Archives of Brazil and is Director of the InterPARES 3 Project's TEAM Brazil.

---

[19] Duchein, "Os arquivos na torre de Babel," ibid., 22. The citation was translated to English by the author of this paper.