

White paper.

Communicating the Text Encoding Initiative to a Multilingual User Community.

del Rio Riande, Gimena y Alles Torrent,
Susanna.

Cita:

del Rio Riande, Gimena y Alles Torrent, Susanna (2023). *Communicating the Text Encoding Initiative to a Multilingual User Community*. White paper.

Dirección estable: <https://www.aacademica.org/gimena.delrio.riande/212>

ARK: <https://n2t.net/ark:/13683/pdea/Z6Y>



Esta obra está bajo una licencia de Creative Commons.
Para ver una copia de esta licencia, visite
<https://creativecommons.org/licenses/by-nc-nd/4.0/deed.es>.

Acta Académica es un proyecto académico sin fines de lucro enmarcado en la iniciativa de acceso abierto. Acta Académica fue creado para facilitar a investigadores de todo el mundo el compartir su producción académica. Para crear un perfil gratuitamente o acceder a otros trabajos visite: <https://www.aacademica.org>.

**Communicating the Text Encoding Initiative
to a Multilingual User Community
Report**

September 2022

By

Susanna Allés-Torrent

&

Gimena del Rio Riande

1. The Project *Communicating the Text Encoding Initiative to a Multilingual User Community*

The Text Encoding Initiative was, from its outset, very much a Western, English-language effort. With early centers of support at such places as Oxford University, the University of Virginia, and Brown University, it could hardly be otherwise. Its remit, however, is global. Primary source documents written in languages as diverse as Chinese, Mayan, Coptic, Japanese or Arabic, among others, are published in TEI. The Guidelines “are addressed to anyone who works with any kind of textual resource in digital form”¹ and they represent a major and long-lived contribution to the continuously developed infrastructure of digital scholarship.²

The TEI Guidelines are written and edited entirely in English, with (incomplete) translations into several languages. The efforts to internationalize the TEI's documentation date back at least to 2005. An initiative led by the late Sebastian Rahtz developed a first infrastructure to support translations and solicited community efforts to provide them. This effort resulted in partial translations in Japanese, Chinese, Korean, Italian, Spanish, French, and German. Over the years since that initial effort, periodic updates have been made to individual languages, most recently German, Japanese, and Spanish.

The project *Communicating the Text Encoding Initiative to a Multilingual User Community* (Mellon Reference No.: 2001-07353) was funded by the Andrew W. Mellon Foundation through the program Public Humanities³. It ran throughout a year beginning in March 2021, and was subsequently extended until December 2022. The main goals of this project are threefold and related to the support of the TEI internationalization group supported by the same Consortium⁴: first, the development of a new infrastructure, translate.tei-c.org, to improve the user experience of multilingual translation of the TEI Specs⁵; second, a survey on

¹ <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/AB.html>.

² In 2017 the TEI Consortium was awarded the Zampolli Prize by the Alliance of Digital Humanities Organizations in honor of its 30th anniversary: <https://adho.org/announcements/2017/2017-adho-awards>.

³ See grant details at: <https://mellon.org/grants/grants-database/grants/duke-university/2001-07353/>

⁴ More information about this working grupo is available at: <https://tei-c.org/activities/workgroups/internationalization-j18n-workgroup/>. Also, the foundations of this Mellon project might be found at the TriangleSci event organized in 2019:

<https://trianglesci.org/2019/07/17/communicating-the-tei-to-a-multilingual-user-community/>.

⁵ The platform was developed by Hugh Cayless (Duke University) and is available at: https://translate.tei-c.org/doc_en.html.

the use of the TEI among the Hispanophone community, and, third, the development of TEI training materials —tutorials, exercises, examples and academic resources— in Spanish.

This report focuses mainly on the survey that the team conducted during the grant period, in Spring 2022, where approximately 130 TEI users offered valuable insights concerning the adoption and use of the TEI, and the needs of this community. First, however, we will give some historical background and a general overview of the state of the art of the TEI in Spanish. Then we will examine the current challenges and opportunities for adopting this standard of digital encoding and publishing texts in the Humanities and Social Sciences. Finally, we will highlight the work that has been done in creating new educational materials and resources for learning the TEI, as well as the many activities we have organized to broadcast its use and create community.

2. State of the art of the Text Encoding Initiative in the Hispanophone World

The digital humanities (DH) field has experienced unstoppable global growth over the last ten years. The methodologies, methods, resources and tools developed by the Text Encoding Initiative (TEI) and scholars from its community have become part of the DH commons all over the globe. As Susan Hockey had made clear in her “The History of Humanities Computing” in the first *Companion to Digital Humanities*:

If one humanities computing activity is to be highlighted above all others, in my view it must be the TEI. It represents the most significant intellectual advances that have been made in our area, and has influenced the markup community as a whole (Hockey 2004: XX).

This is especially true for the English-speaking world, where standards for creating searchable, enriched, interoperable and well preserved texts and databases were part of early projects. Still, in the last five years, one can observe how DH, as a discipline, is in the midst of a significant *global turn* that has come to question many aspects of our linguistic and technical practices, underlining barriers to the participation of a globalized population and other equity-seeking groups within the field (O’Donnell 2012; Galina 2013; Fiormonte 2014; O’Donnell et al. 2016a; O’Donnell et al. 2016b; Fiormonte and del Rio Riande 2017). As the use and study of computation in the Humanities grows, its core literature is beginning to incorporate increasing amounts of work by researchers from other disciplines and from networks and regions outside the English-speaking Academies in the Global North (del Rio Riande 2015a, 2015b, 2016a, 2016b). This same applies to TEI practices within the global community. Next we will look at where TEI has appeared as part of Hispanophone workshops, courses, projects, and/or teaching resources.

2.1. DH associations in Spain and Latin America, and TEI training

Despite the early interest in the intersection between Humanities and the use of technology (Marcos Marín 1986; del Rio Riande 2015, 2016a), only in recent years has the DH landscape become recognizable as a field in terms of Hispanophone communities. Several DH associations

in Latin America and Spain have emerged since 2012: the very first was the *RedHD* in Mexico⁶, followed by the *Associação das Humanidades Digitais* (AHDig)⁷ in Brazil, and the *Sociedad Internacional de Humanidades Digitales Hispánicas* (HDH) in Spain⁸. In Argentina, the *Asociación Argentina de Humanidades Digitales* (AAHD)⁹ was created in 2013, during THATCamp Buenos Aires¹⁰. More recently, the *Associació d'Humanitats Digitals Catalanes* (AHDCat)¹¹ was created with the intention to feature Catalan linguistic and cultural projects, and, in Latin America, the *Red Colombiana de Humanidades Digitales*¹² have also come into play, as well as some other initiatives in Perú and Uruguay¹³.

RedHD and HDH were the first associations that included TEI workshops in their first Conferences in 2012¹⁴. Both associations have continued offering informal half-day or one-day TEI training in their events. Also, in 2013 the first TEI workshop was taught in Argentina as part of the aforementioned THATCamp, and a four-hour workshop was held at the first National Conference in 2014 and continued in 2016's conference as part of workshops and panels¹⁵. Both authors of this report participated at these events.

In this last year, although COVID has hindered these kinds of in person activities, most of these associations have continued to organize annual meetings or international conferences with sessions devoted to the TEI¹⁶.

⁶ Red de Humanidades Digitales (RedHD), www.humanidadesdigitales.net.

⁷ Associação das Humanidades Digitais (AHDig), <https://ahdig.org/>.

⁸ Humanidades Digitales Hispánicas. Sociedad (HDH), <http://www.humanidadesdigitales.org/inicio.htm>.

⁹ Asociación Argentina de Humanidades Digitales (AAHD), <http://aahd.net.ar/>.

¹⁰ For this event, see <http://buenosaires2013.thatcamp.org/>.

¹¹ Associació d'Humanitats Digitals Catalanes, <https://ahd.cat/>.

¹² Red Colombiana de Humanidades Digitales, <http://www.rehdi.co/>.

¹³ *Asociación Uruguaya de Humanidades Digitales* (HDU), <https://www.facebook.com/DHUruguay/>. This institutional growth of associations is also a European trend, that in the past years saw the creation of the *Associazione di Informatica Umanistica* (AIUCD), <http://www.aiucd.it/>, in Italy, the *Digital Humanities in Deutschsprachigen Raum* (DHD), <https://dig-hum.de/>, in Germany, and *Humanistica-Association francophone des humanités numériques/digitales*, <http://www.humanisti.ca/>, created as a Francophone association in which it is the linguistic *liason* that prevails, rather than the geographic element. All of them count with listserv as their main channel of communication to spread the news.

¹⁴ See <http://humanidadesdigitales.net/1er-encuentro-de-humanistas-digitales/> and <https://humanidadesdigitaleshispanicas.es/i-congreso-internacional-de-la-hdh/>.

¹⁵ See <https://www.aacademica.org/aaahd2014/registerDocuments/1.pdf> and <https://www.aacademica.org/aaahd2018/startDocuments/4.pdf>.

¹⁶ As an example, see the programs at the V Conference of the HDH (Spain), <https://hdh2021.org/> or the V International Conference of the AAHD (Argentina), <https://www.aacademica.org/aaahd2022>.

2.2. Digital Scholarly Editions Projects

One of the main uses of the TEI is the creation of digital scholarly editorial projects, although its use for corpus creation has also now been well adopted. In the case of Spain, the interest in the use of TEI for Digital Scholarly Editions (DSE) started in early days (Marcos Marín 1986, 1994), but scientific literature about the subject was sparse in this country until 2008 (Lucía Megías 2008). Some general overviews that concentrate just on Spain have been published, but they do not reflect any research about the TEI in Latin America (Spence 2014; Spence and González-Blanco 2014; Allés-Torrent 2017) or they only cover case studies of projects (Revenge 2014; Rojas Castro 2017). Publications dealing with TEI and its use in Spanish and Latin American projects have started growing since 2015 (Faulhaber 1996; Fradejas 2009; Allés-Torrent 2015, 2017, 2020; González-Blanco et al. 2014, 2015; del Rio Riande and Zubillaga 2015; Priani 2017, among others)¹⁷.

One of our lines of work during the grant period consisted of inventorying projects dealing with Spanish texts¹⁸. More than forty examples of very different nature show the rise of the use of the TEI and projects' willingness to share their encoded work. What this list demonstrates is that the landscape of projects related to Hispanic texts using TEI is growing steadily. The DSEs which have been published deserve special attention, and show some very interesting features. However, their coverage in the two existing catalogs created by Greta Franzini and Patrick Sahle¹⁹ is far from satisfactory compared with other languages such as English, Latin, French, German or Italian. In this regard, there are some facts that we would like to bring forward. On one hand, DSEs are more common among scholars interested in Medieval texts (such as *Cantar de Mio Cid*)²⁰ and Early Modern Literature—the so-called Golden Age period—for which we find editions of classical plays such as Miguel de Cervantes' *La Entretenida*²¹ or Lope de Vega's *La Dama Boba*²², and others on *Don Quixote*, largely through

¹⁷ In our project, *Text Technologies Hub* (TTHub) we offer an exhaustive list of publications regarding TEI in Spanish, see <https://tthub.io/recursos#section-bibliografia-en-zotero>

¹⁸ The list can be consulted here <https://tthub.io/proyectos-que-utilizan-xml-tei>

¹⁹ A *Catalogue of Digital Editions*, by Greta Franzin, <https://dig-ed-cat.acdh.oeaw.ac.at/>, offers a total of 320 projects of which only 19 are related to Spanish, and the *Catalogue of Digital Scholarly Editions*, by Patrick Sahle, <http://www.digitale-edition.de/> has a total of 855 of which only 18 are Spanish related.

²⁰ See: <https://miocid.wlu.edu/?v=nor>.

²¹ <http://entretenida.outofthewings.org/index.html>

²² <http://damaboba.unibo.it/index.html>

funded projects (Allés-Torrent 2017)²³. But early Mexican manuscripts are also represented with *El sitio de Guaman Poma* and the *Codice Mendoza*²⁴. Other initiatives deal with the 19th and 20th century, such as the digital edition *En el ojo del huracán. Cartas de Yltramara a España, 1823*²⁵, *The Pérez Galdós Edition Project*²⁶, the *Manuscrito digital de Juan Goytisolo*²⁷ or *Obras Completas de José Luis Romero*²⁸, among others. It is interesting to note that many of them were carried out in non-hispanic institutions (*Cantar de Mio Cid*, *El sitio de Guaman Poma*, *Manuscrito digital de Juan Goytisolo*), and that others, like the *Quijote Interactivo*, a great site funded by the Biblioteca Nacional de España, were built with technologies such as Adobe Flash Player, or not using textual markup, as in the case of the *Códice Mendoza*, funded by the INAH in Mexico, which seems just to be focused on the images of the mesoamerican manuscript. On the other hand, even when they used the TEI, few of these editions declare their use of it, and examples of DSEs offering their encoded texts remained rare, such as *La Dama Boba*, or *Biblioteca Digital del Pensamiento NovoHispano*²⁹ (Priani 2017).

As for corpus creation, TEI suffered from discontinued use in the two big Spanish projects in digital libraries, the *Biblioteca Virtual Miguel de Cervantes*³⁰ and *Corpus Diacrónico del Español* (CORDE).³¹ However, there is a new revival for the adoption of TEI through initiatives as CHARTA,³² and other outstanding projects, such as P.S. Post Scriptum, *A Digital Archive of Ordinary Writing (Early Modern Portugal and Spain)*³³. Most of these rely on the TEITOK platform.³⁴

Two other more recent trends that we have detected are, firstly, the adoption, especially in Latin America, of Minimal Computing approaches that have started co-existing with TEI DSEs, like the editions of HD Lab, at CONICET (Argentina)³⁵, which have also inspired projects in

²³ See for example, Electronic Variorum Edition of the Quixote, <http://www.csd.tamu.edu:8080/veri/index-en.html>

²⁴ For *Guaman Poma*, see: <http://www5.kb.dk/permalink/2006/poma/info/en/frontpage.htm> And for the *Codice Mendoza*, <http://codicemendoza.inah.gob.mx/inicio.php>.

²⁵ <http://www.cartas-de-ultramara.net/>

²⁶ <https://www.dhi.ac.uk/galdos/>

²⁷ <http://goytisolo.unibe.ch/>

²⁸ <https://jlromero.com.ar/>

²⁹ <http://bdpnh.unam.mx>

³⁰ <https://www.cervantesvirtual.com/>

³¹ <http://corpus.rae.es/cordenet.html>

³² See this umbrella project for old hispanic documents, <https://www.redcharta.es/>

³³ See the corpus at <http://teitok.clul.ul.pt/postscriptum/index.php>

³⁴ Available at <http://www.teitok.org/index.php>

³⁵ See: <http://hdlab.space/biblioteca-digital/> For the Minimal Computing approach, see <https://go-dh.github.io/mincomp/>

Colombia.³⁶ Second, the US Hispanic community, the largest minority in the US, is demonstrating some impressive digital initiatives and projects, such as the *Relaciones Geográficas* (providing data for 71 towns in New Spain)³⁷ and the first US Digital Humanities Center for Latina/o Studies at the University of Houston and their editorial work at Arte Publico Press.³⁸

2.3. TEI training: Informal and Formal Education

As far as DH training, there has been informal training conducted since the 2000s, mainly workshops, devoted to short introductions to the basics of XML-TEI. These events have been more frequent in the last ten years and especially in Spain.³⁹ One of the most typical venues for training are the workshops preceding DH conferences organized by the different associations.

During the early 2000s, there were some unsuccessful attempts to offer graduate DH certificates in Spain, such as at the one of the Escuela Superior de Ingeniería Informática at the Universidad de Castilla-La Mancha. However, in the last years, some universities have tried to join forces merging different disciplines, such as the Universidad de Salamanca, which offers a MA in Textual Heritage and Digital Humanities⁴⁰, and the Universidad Pablo Olavide (Seville) that does the same in a MA in History and Digital Humanities⁴¹. Also, the Universitat Pompeu Fabra and the Universitat Autònoma de Barcelona offer a joint MA under the label “Humanitats i Patrimoni Digitals”⁴² where a course on TEI is offered by Ramón Valdés. Regarding Digital Literature, the Research Group LEETHI has been organizing a Digital Literature postgraduate course at the Universidad Complutense de Madrid, and also the University of Barcelona⁴³. In the

³⁶ See: <https://suarezg106.gitlab.io/conversacion-cafit/>.

³⁷ See: <https://www.arcgis.com/apps/Cascade/index.html?appid=1fcabf740a844d9d80d5bf0248416f47>.

³⁸ See <https://artepublicopress.com/digital-humanities/>

³⁹ Take into consideration, for example, the trainings offered by Alex Bia Plata, <https://alexbia.umh.es/docencia/talleres/> since the early 2000 until nowadays. Another example, organized by one of the authors of this report, was “XML-TEI for Ancient and Medieval Lexicographical Works”, celebrated at the Institució Milà i Fontanals, CSIC- Barcelona in 2013, <http://gmlc.imf.csic.es/2013/Workshop/index.php>.

⁴⁰ See Máster Universitario en Patrimonio Textual y Humanidades Digitales: <http://www.usal.es/patrimonio-textual-y-humanidades-digitales>.

⁴¹ See Máster Universitario Historia y Humanidades Digitales: <https://www.upo.es/postgrado/Master-Oficial-Historia-y-Humanidades-Digitales>.

⁴² See: <https://www.uab.cat/web/estudiar/l-oferta-de-masters-oficials/informacio-general/humanitats-i-patrimoni-digital-1096480139517.html?param1=1345803179474>

⁴³ See: <https://www.il3.ub.edu/es/postgrado/postgrado-literatura-comparada-literatura-digital.html>.

realm of online education, the Laboratorio de Innovación en Humanidades Digitales (LINHD), at the Universidad Nacional de Educación a Distancia (UNED) is the only institution so far offering a continuous variety of academic DH courses since 2014 in which professors from Spain and Latin America teach (González-Blanco et al. 2017a) in a MA in Digital Humanities.⁴⁴ The Universidad Internacional de La Rioja also offers an online degree, Máster Universitario en Humanidades Digitales.⁴⁵

On the other side of the world, DH training in Latin America can be described in terms of MA degree offered by the Universidad de Los Andes (Colombia)⁴⁶, another offered at the Universidad de Sor Juana Inés de la Cruz in México, and a six-month postgraduate course offered in Argentina at Universidad de Ciencias Empresariales y Sociales (UCES)⁴⁷. Only in the last case is a complete TEI course is taught. There are many initiatives worth highlighting, such as the Seminario HD in México at UNAM (Universidad Nacional Autónoma de México, Facultad de Filosofía y Letras)⁴⁸ (Priani 2007), and the many DH events organized by HD Lab (CAICYT-CONICET),⁴⁹ along with many others held at the Universidad de Buenos Aires⁵⁰, Universidad de La Plata,⁵¹ and Universidad Nacional de Mar del Plata in Argentina, at Pontificia Universidad Católica in Chile,⁵² and at different events in Cuba⁵³.

Overall, we can conclude by saying that alongside informal training which is occasional and frequently lacks continuity, most formal offerings at the postgraduate level include TEI courses. This panorama will improve with the growth of projects using text encoding practices and the availability of resources to learn it.

⁴⁴ See: <https://linhd.uned.es/cursos-y-formacion>

⁴⁵ See: <https://www.unir.net/humanidades/master-humanidades-digitales/>

⁴⁶ See: <https://tinyurl.com/y8rb2reu>

⁴⁷ See:

<https://www.uces.edu.ar/carreras-escuela-negocios/gestion-del-talento-humano/diplomatura-humanidades-digitales>.

⁴⁸ See: <http://www.humanidadesdigitales.net/seminario-hd/>.

⁴⁹ See: <http://www.caicyt-conicet.gov.ar/micrositios/hd/>

⁵⁰ See the first DH four-month course taught by Gimena del Rio Riande at Facultad de Filosofía y Letras in 2015: <https://tinyurl.com/y8rb2reu> and others as part of the Medieval Hispanic Literature classes (2017).

⁵¹ An online two-month course taught by Gimena del Rio and Natalia Corbellini (2016) in an Argentinian university. See: <https://tinyurl.com/y8rb2reu>.

⁵² See: <https://tinyurl.com/y8eryh7d>.

⁵³ See: <http://www.caicyt-conicet.gov.ar/micrositios/hd/?p=905>, workshop by Gimena del Rio Riande and Elena González Blanco on the use of TEI for Digital Scholarly Editions (2017), <http://www.caicyt-conicet.gov.ar/micrositios/hd/?p=1150>, and two workshops devoted to the Digital Editing inside and outside Academy, with special emphasis on the use of TEI.

2.4. TEI training materials in Spanish: TThub

Many debates on the anglo-US centrism as well as on the heterogeneity of approaches to DH are taking place. These include how to understand DH, what to teach, and, essentially, whether *digital humanities* equals *humanidades digitales* (del Rio Riande, 2015, 2016ab). What seems obvious is that all global practices are born out of local contexts that shape the way we navigate the different levels, locally to globally, and vice versa.

In the case of TEI learning, *TEI By Example* was developed several years ago, providing explanations, online tests, and exercises, albeit only in English. Its creators, Van den Branden, Terras, and Vanhoutte stressed the need for a “more user friendly, comprehensive, and interactive than the online workshop materials which are currently presented as stand alone teaching materials” (Terras, Van Den Branden, and Vanhoutte 2009: 299). After all these years, while some of the issues that they emphasized remain the same, the success of this resource has been proved extensively and has significantly improved the overall situation for anglophone TEI users. Recently, other interesting initiatives offering pedagogical materials have emerged, such as #dariahTeach.⁵⁴ This European project has inaugurated a collection of courses, three of which deal with TEI: “Text Encoding and the TEI” by Susan Schreibman and Roman Bleier⁵⁵ in English with French and Spanish versions, and “Digital Scholarly Editions: Manuscripts, Texts and TEI Encoding” by Elena Pierazzo and Marjorie Burghart.⁵⁶ These three courses, which have explanatory videos, detailed contents, and exercises and that can be considered pioneering work in the field of DH online teaching, were written originally in English. And, although for some of them there are translations, the divide persists as each language and each social, cultural and academic context have specific needs and require specific examples taken from the local praxis (Isasi & Rojas Castro 2021).

Therefore, it is imperative for the Spanish-speaking community to make available more reliable resources, following the models of *TEI By Example* and #dariahteach. Responding to this need, and funded, as part of this grant, our team further developed the project *Text Technologies Hub* or TThub.⁵⁷ TThub was created in 2019 and its principal aim is to function as an open

⁵⁴ See: <https://teach.dariah.eu/>

⁵⁵ See: <https://teach.dariah.eu/course/view.php?id=23>

⁵⁶ See: <https://teach.dariah.eu/course/view.php?id=32>

⁵⁷ See: <http://tthub.io>

access hub of text technologies and digital scholarly edition resources in Spanish, conceived especially for the Hispanophone community. Our goal is to serve as a hub of available online materials, resources, news, softwares and technologies that can potentially serve those interested in textual studies, digital editing, corpus construction, and digitization processes in Spanish.

The site is divided into two main sections: *Aprende* (Learn) and *Recursos* (Resources). In the section *Aprende* we focus on offering different tutorials that provide an independent learning curve, as Spence and Brandao have suggested

(...) are strongly influenced by the self-directed learning model of *online tutorials* popular for learning to use digital tools, a genre that may be text based or video based and which typically employs practice-based ‘learning by example’ to achieve mastery of a given programme or tool. (Spence & Brandao 2020)

We have made available different situated tutorials that are not translations but original materials with examples and exercises drawn from Spanish literature and historical texts. Due to limited space, in this report, we will only highlight two of them: 1) a version of the teaching materials inspired by the several iterations of the Master degree in DH created in January 2014 at the LINHD-UNED. The course is presented under the name *Introducción a la Text Encoding Initiative* and already enjoys of certain popularity among the Hispanophone community,⁵⁸ 2) a new and more complete tutorial, written during 2022, *Edición y publicación de textos con TEI*.⁵⁹ This tutorial deals with deeper TEI encoding and publishing methodologies and practices.

As for our web infrastructure and workflow, all our materials are hosted in GitHub and from there, files are loaded to be rendered and embedded into the current site by a PHP script. This infrastructure allows us to harvest any materials that have the same format (e.g. written in markdown and hosted in GitHub). Hence, the possibility to import other tutorials such as the ones published by the *Programming Historian*⁶⁰. Following a new interface design of the site, all tutorials have the same look and feel. We believe resources like these will facilitate the learning experience of scholars and students in a self-taught experience, as part of the materials of an academic syllabus, or even as the common ground for teaching at workshops and other training events.

⁵⁸ Available at <https://tthub.io/aprende/tutorial/introduccion-text-encoding-initiative>

⁵⁹ See: <https://tthub.io/aprende/tutorial/edicion-y-publicacion-textos-tei/>

⁶⁰ See for example: <https://tthub.io/aprende/tutorial/publicacion-web-archivos-tei-ceteicean>

APRENDE RECURSOS

TTHub
RECURSOS SOBRE TECNOLOGÍAS DEL TEXTO Y EDICIÓN DIGITAL
ACERCA DE TTHUB PARTICIPA

Introducción a la Text Encoding Initiative

- 1 Definición, aplicaciones prácticas y recursos
- 2 El lenguaje XML
- 3 Estructura básica y elementos comunes
- Básicos
- Encabezado (teiHeader)
- Cuerpo del documento
- Divisiones
- Elementos comunes de todo documento TEI
- Bibliografía
- 4 Guías directrices de la TE
- 5 Tipologías textuales
- 6 Esquemas RNG

INICIO / APRENDE / INTRODUCCIÓN A LA TEXT ENCODING INITIATIVE

3 Estructura básica y elementos comunes de los documentos XML-TEI

Cuerpo del documento

El cuerpo del documento corresponde al elemento `<text>` que puede contener, a su vez, tres (sub)elementos:

- `<front>`: se utiliza para marcar elementos paratextuales que preceden el texto, tales como prefacios, prólogos, cartas dedicatorias, una lista de personajes, etc. y no es obligatorio.
- `<body>`: se consagra al texto propiamente dicho del documento y es obligatorio.
- `<back>`: puede utilizarse para añadir los apéndices, índices, cronologías, bibliografías, etc. No es obligatorio.

Siguiendo con el ejemplo anterior podríamos marcar el texto con la siguiente estructura:

```

<text>
  <front>
    <div type="prologo">
    </div>
  </front>
  <body>
    <div type="tomo" n="1">
      <div type="capitulo" n="1"> ... </div>
      <div type="capitulo" n="2"> ... </div>
      <div type="capitulo" n="3"> ... </div>
    </div>
    <div type="tomo" n="2">
      <div type="capitulo" n="1"> ... </div>
      <div type="capitulo" n="2"> ... </div>
      <div type="capitulo" n="3"> ... </div>
    </div>
  </body>
</text>

```

En línea de máxima, como sucede con los nombres de los elementos y de los atributos, es mejor no utilizar acentos ni espacios tampoco en los valores de los atributos.

Cita

Allés Torrent, Susanna (2019). "Introducción a la Text Encoding Initiative". *TTHub. Text Technologies Hub: Recursos sobre tecnologías del texto y edición digital*. <https://TTHub.io/aprende/introduccion-a-tei/>

Copiar

Figure 1. Design for the tutorials in TTHub. <https://tthub.io/>

In *Recurros*⁶¹ we present different repositories featuring examples of projects that provide their TEI files in open access, as well as a searchable list of tools, an exhaustive bibliography, a Zenodo community where to upload materials, and other resources such as catalogs of digital editions, best practices guidelines, and direct access to the new TEI translation platform.

⁶¹ Available at: <https://tthub.io/recurros>

Inicio / Recursos

Proyectos que utilizan XML-TEI

Este repositorio ofrece una lista de proyectos que utilizan una codificación en XML-TEI. Todos los proyectos tienen en común que trabajan con textos en español y que comparten sus archivos digitalizados, la mayoría a través de repositorios en TTHub. Los ejemplos pueden clasificarse según su tipología textual (poesía, novela, crónica, teatro, etc.) y ofrecen un enlace para ir a la web del proyecto como a su código fuente. Tienen a disposición ejemplos reales de codificación es sumamente útil que sirve como modelo a la hora de marcar nuestros textos. Podrán otros recursos donde consultar más ejemplos en otras lenguas, como el [TEIHub](#). Esta lista se obtiene de una hoja de cálculo en Google Drive a través de un script PHP. ¿Conoce otros proyectos que utilizan XML-TEI para codificar sus textos? Pues mándanos la referencia a través de la sección [Participa](#).

Filtrar por idioma

[Español](#)
[Catalán](#)
[Gallego](#)
[Castellano](#)
[Catalán](#)
[Español](#)
[Español](#)
[Español](#)
[Español](#)
[Español](#)
[Español](#)
[Español](#)

Página 1: Mostrando 20 de 43 resultados

<p>7 Partidas Digital</p> <p>Codificación en XML-TEI de los textos romances de las Siete Partidas</p>	<p>Edición: José María López de Letamendi Asesor: Alfonso Xalón</p>	<p>TEIHub</p> <p>Proyecto</p> <p>Repositorio</p>
<p>AB34RD</p> <p>Texto de parlamento español de siglo XVII. La última actualización de repositorios del año 2014 y no ofrece un realme file con datos relacionados.</p>	<p>Edición: Text Creation Partnership Asesor: Valia</p>	<p>TEIHub</p> <p>Proyecto</p> <p>Repositorio</p>
<p>Apariencias</p> <p>Repositorio de archivos TEI de teatro español para tesis sobre teatro de Siglo de Oro</p>	<p>Edición: Héctor Ruiz Soto Asesor: Lore de Vega</p>	<p>TEIHub</p> <p>Proyecto</p> <p>Repositorio</p>
<p>BDPN</p> <p>Ediciones críticas de textos impresos y manuscritos editados en Nueva España durante el siglo XVII</p>	<p>Edición: Proyecto BDPN/UNAM Asesor: Varios</p>	<p>TEIHub</p> <p>Proyecto</p> <p>Repositorio</p>
<p>Becerro Galicano de San Millán de la Cogolla</p> <p>Ejemplo de codificación de un cartulario medieval</p>	<p>Edición: Varios Asesor: Varios</p>	<p>TEIHub</p> <p>Proyecto</p> <p>Repositorio</p>
<p>BETTE</p> <p>La Biblioteca Electrónica de Textos del Teatro en Español (BETTE) es una colección de textos teatrales de la Edad de Plata española, codificada en XML-TEI por el grupo GHE2I (obra ANBAUN/R) de la Universidad Internacional de Aragón (UNIR)</p>	<p>Edición: IMA, IIA, IRIPIPIR Asesor: García Llorca, Valle Indán, Muñoz Best, Dehagany, Unamuno, Valera, Clarin, Sances y Clazote</p>	<p>TEIHub</p> <p>Proyecto</p> <p>Repositorio</p>
<p>Calderón Drama Corpus (CalDraCor)</p> <p>Archivos TEI de Manuscritos de Calderón Drama Corpus (CalDraCor)</p>	<p>Edición: Institute of Romance Languages and Literatures University of Tübingen Asesor: Pedro Calderón de la Barca</p>	<p>TEIHub</p> <p>Proyecto</p> <p>Repositorio</p>
<p>Castrojeriz</p> <p>Edición digital de Glosas Castellanas of Regimiento de Principes</p>	<p>Edición: Wardhis Gilje-Loverson Asesor: Juan García de Castrojeriz</p>	<p>TEIHub</p> <p>Proyecto</p> <p>Repositorio</p>
<p>CN-HD-repo</p> <p>Repositorio piloto del proyecto Corpus Normativo Hispano (Iniciativa digital) / codificado en TEI (en construcción)</p>	<p>Edición: Victor Goyá Asesor: Varios</p>	<p>TEIHub</p> <p>Proyecto</p> <p>Repositorio</p>

Figure 2: Repository of projects using XML-TEI

Finally, in addition to these two options (Learn and Resources), we have added a section to invite participation through improving our repositories, joining our community, or creating new materials to be published in the TTHub.

3. Our Survey: Access, Use, and Needs for the Hispanophone TEI Community

The bilingual survey “Use of the Text Encoding Initiative in Spanish for Digital Humanities Projects” / “Uso de la Text Encoding Initiative en español para proyectos de Humanidades Digitales” was developed in the last months of 2021 and distributed on January 24, 2022, through different mailing lists.⁶² It remained open until March 10th, 2022. The survey consisted of 22 questions and it was not supposed to take more than 10 minutes to complete. Contributors to the survey remained anonymous, although we collected data related to affiliation and nationality in order to obtain a better understanding of our community from a geographical and geopolitical perspective. We looked for participants who had used the TEI for their research or taken any TEI course at any level. We also encouraged participation by those who were either part of the Global Spanish-speaking community or work with Spanish texts encoded in TEI in any country. This range of participation means we can relate the use of Spanish as a language for academic communication to the language used at participants’ university for learning and or writing/publishing, and to projects that deal with Spanish texts in a Hispanic or non-Hispanic country.

The software used to create and distribute the survey was Qualtrics,⁶³ and the user license was provided by the University of Miami. 134 participants started the survey, though only 107 answered all the sections and questions. 77 of these 107 answers were in Spanish and 28 in English.

⁶² The Survey was distributed in the following listservs: Humanidades Digitales Latinoamericanas latamhd@lists.humanidadesdigitales.net, Humanidades Digitales Hispánicas HDH@LISTSERV.REDIRIS.ES, TEI mailing list TEI-L@LISTSERV.BROWN.EDU, Digital Medievalist dm-l@uleth.ca, Medieval Texts - Philology Codicology and Technology medtextl@lists.illinois.edu, Asociación Argentina de Humanidades Digitales aahd-l@uleth.ca, RedHD; LLAAR, Humanist, Air-L, UVic, UCarolina; BBWAG, Digital Classicist DIGITALCLASSICIST@JISCMAIL.AC.UK

⁶³ See: <https://www.qualtrics.com>

The survey was divided into four main blocks: first, general information; second, scholarly training and the TEI; third, use of the TEI; and fourth, teaching and learning the TEI. In what follows, we offer a summary of the results.

3.1. General Information about Participants

This section helped us to find out who is the population nowadays working in TEI projects and where. Also, it gave us information about global exchanges and academic profiles. As we have already mentioned, we offered the survey in Spanish and English, however, as expected, most of the answers were completed in Spanish:

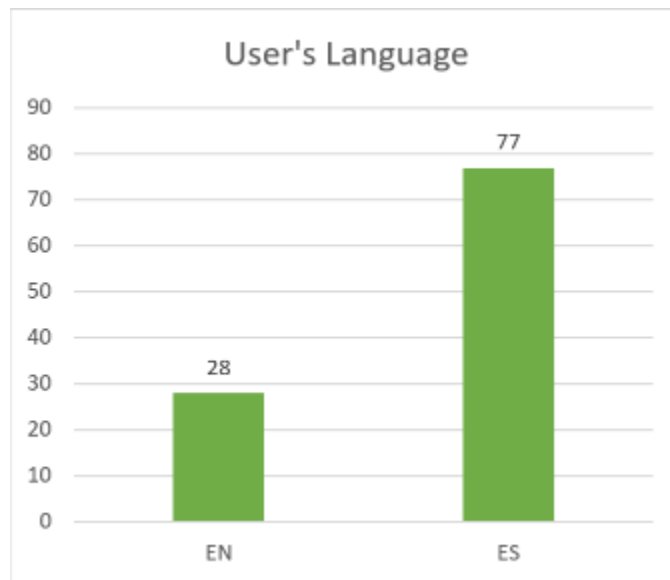


Figure 3. Language used to answer the survey.

According to the answers, most of the people who answered the survey are nowadays living in Spain. This is a big first group of participants, but we can also spot two other groups: one formed by participants living in México and the USA (14 participants in each case) and a third smaller group constituted by participants living in Argentina and Colombia.

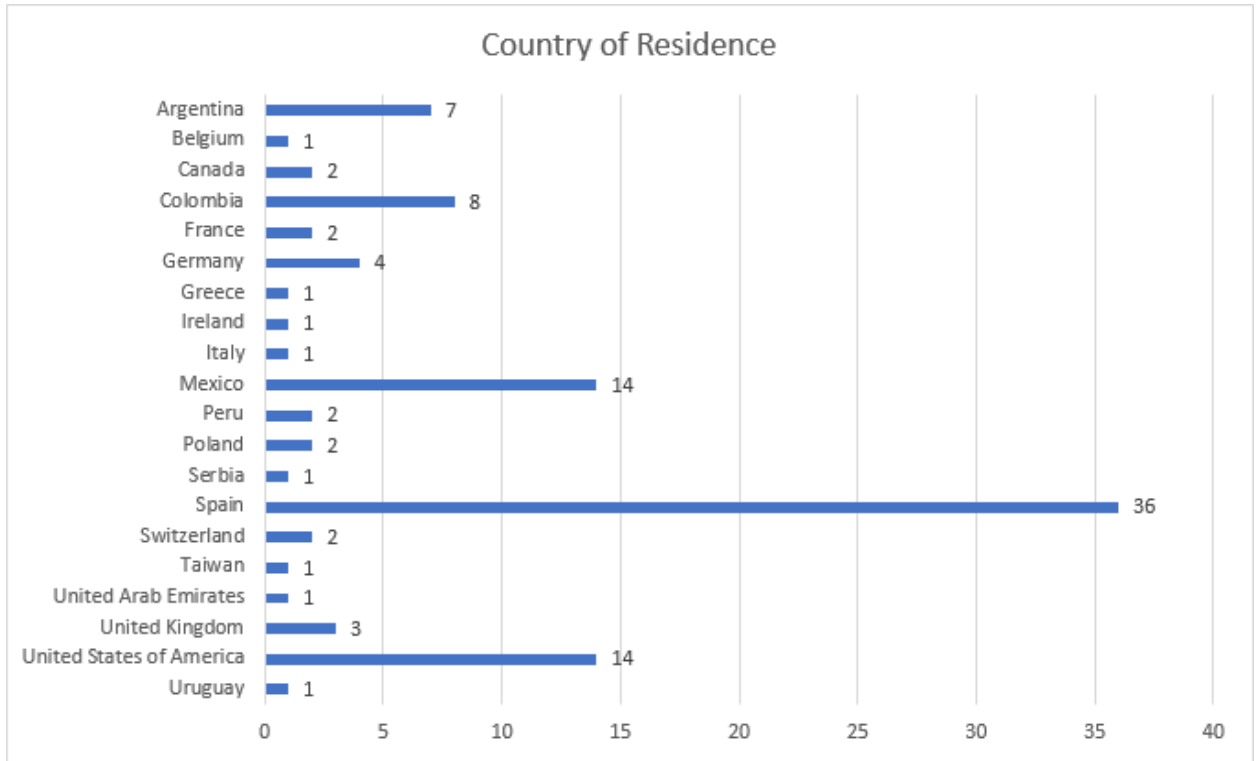


Figure 4. Country where the participants are currently living.

Many participants were born in a different country than the one in which they now live. The majority of these participants are from Spain (42 versus 36):

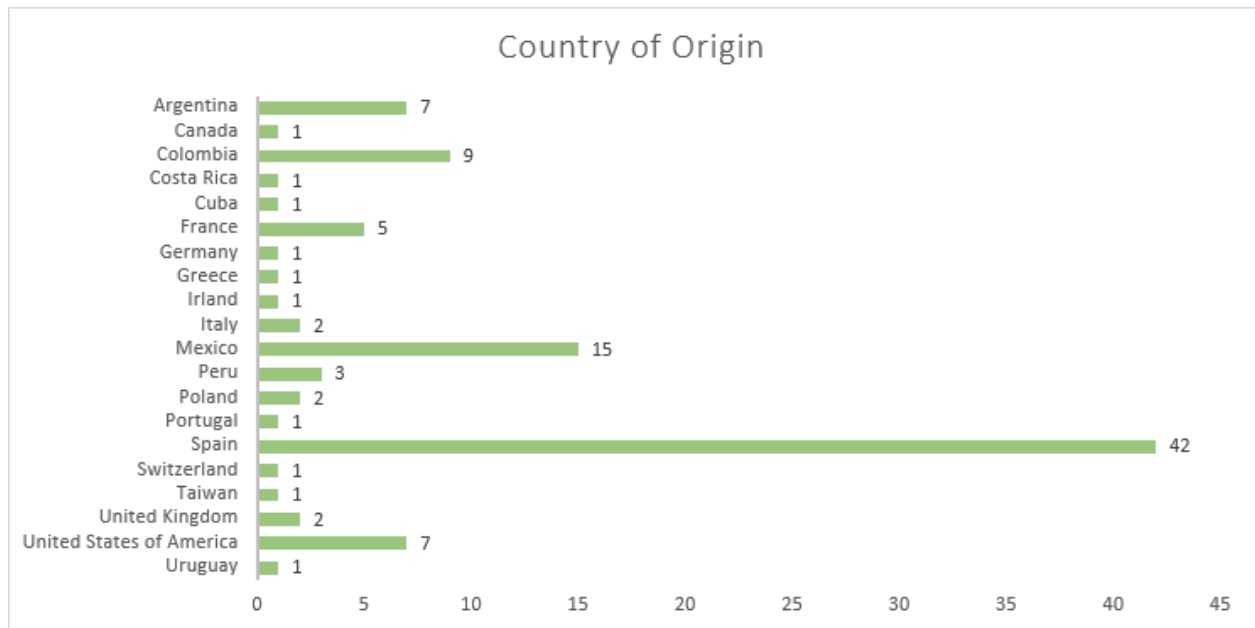


Figure 5. Country of origin of the participants.

The question on the participants' academic position led us to some interesting numbers. Researchers, Associate Professor, and Full Professor seem to be the academic positions of the majority, as a tenured-track group, but not many more than Lecturers and Assistant Professors. MA or Graduate are more popular among the student group rather than postdocs. Also, we find Librarians and only one developer. In the "Other" category, participants responded with categories that could have been matched with one of the proposed ones: PhD student (2), postdoc (1), Research support technician (4), Researcher (2), Archivist (2), Metadata Encoding Specialist (1), Undergraduate student (2), Editor (2), other categories of Faculty (3), Project Manager (1), showrunner (1).

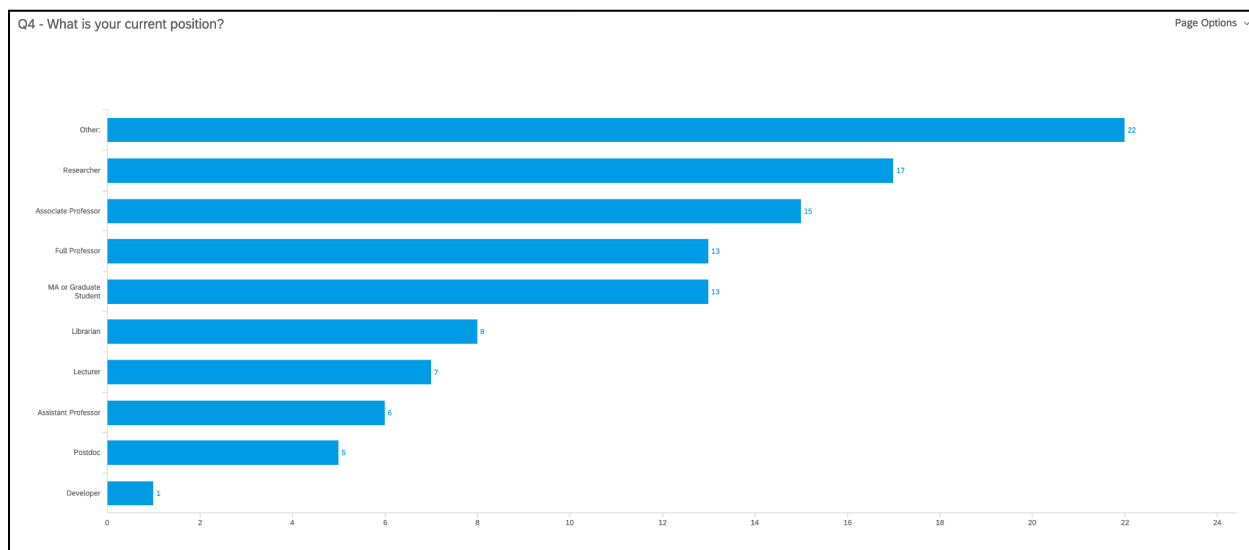


Figure 6. Current position of participants.

The feeling of these responses is, although the scholarly nature of the TEI persists and is represented mainly by academic positions, it is permeating cultural institutions (Libraries, Archives) and being taken in charge also by technical support.

3.2. Scholarly Training and the TEI

The set of questions proposed in the second section, Scholarly Training and the TEI, helped us understand the background of the participants. The answers showed us that some participants had

studied or used the TEI for their Master's (16 answers) or PhD dissertation (16). The smaller percentage was the one related to the use of the TEI in undergraduate studies (6) or as part of librarian support (4). These results are unsurprising as, on the one hand, most of the DH certificates that include TEI training are part of postgraduate courses and, on the other, roles such as DH Librarian are almost non-existent in Spain and Latin America. The category "Other" received the highest score with 59 answers; some participants simply repeated their job (amateur 1, archivist 1, scientific journals 1, Project manager 1, researcher 4, instructor 3), while others specified their current level of studies: Undergraduate (4), Masters degree (3), Graduate level (1). A few participants indicated as well that they took some courses at the university level including Summer Schools (8). But the most striking fact is the number of participants who wrote in "Research Project" (21 answers), that is, people who are involved in the TEI because they are part of an ongoing project.

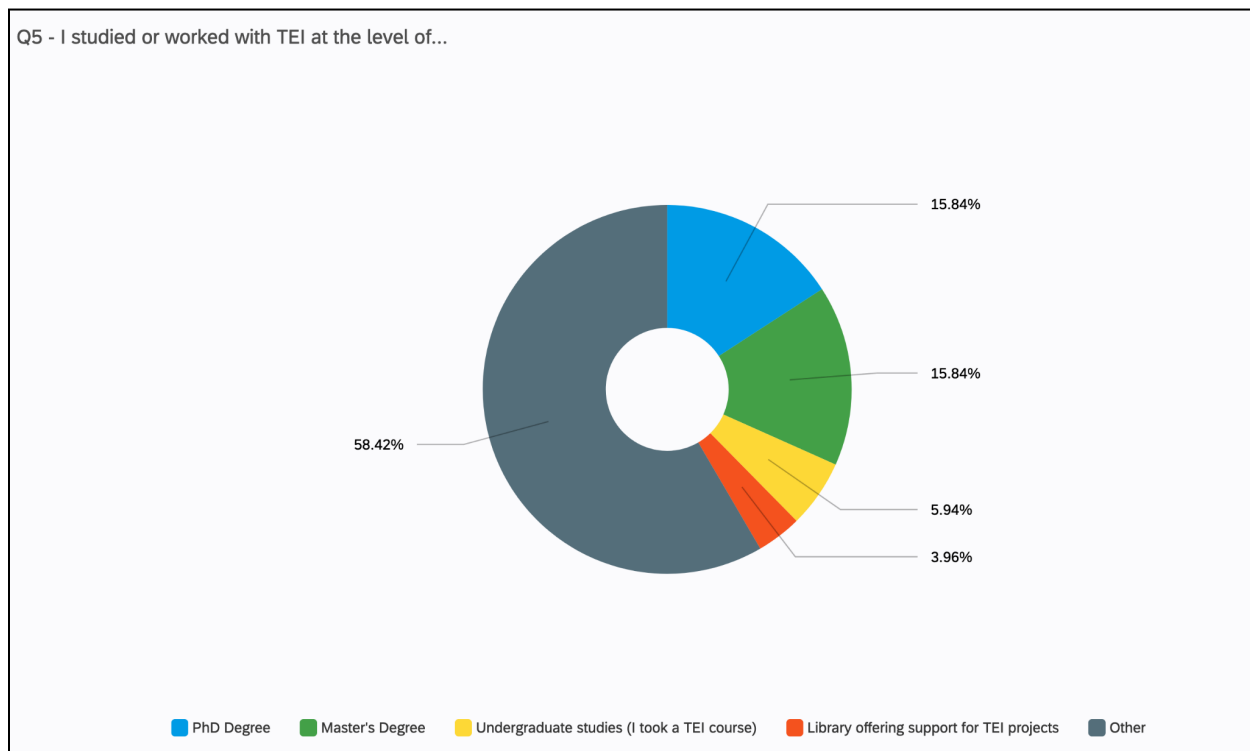


Figure 7. Scholarly training in TEI

Highly interesting as well was to find out that most of the participants had not taken a TEI course as part of their MA or graduate courses. This means that, even though DH offerings are growing, only a few of them provide TEI courses.

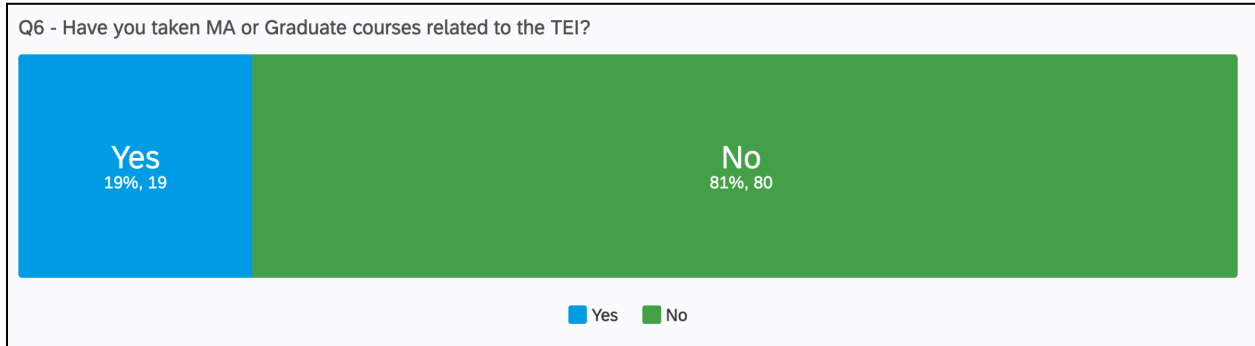


Figure 8. TEI courses taken during MA or Graduate studies

Finally, the percentage of participants who had undertaken their graduate studies in Italy came as a surprise. Mexico too was a leading location. It is difficult to know whether this is a coincidence of the participants' profiles or in fact evidence of a higher training offering at MA and PhD level. Moreover, in most of these cases, the advisor belonged to the same university where the participant had studied, which reinforces the idea of a relatively solid institutional staff and support.

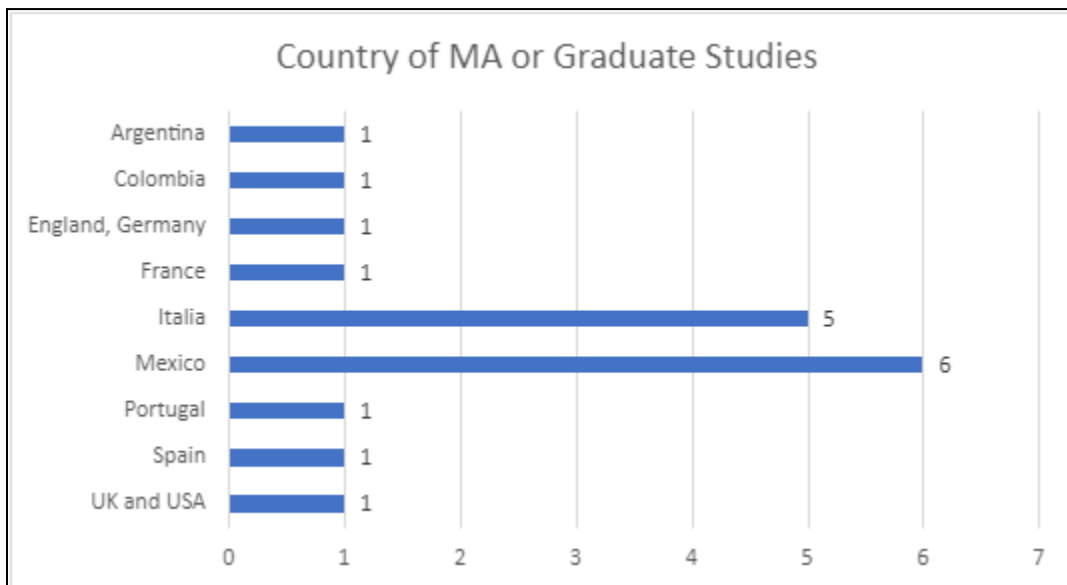


Figure 9. Country of MA or Graduate studies

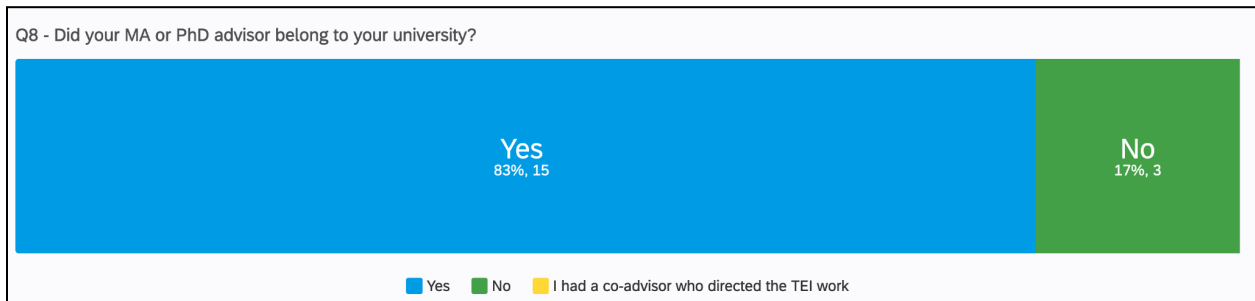


Figure 10. MA or PhD advisors

3.3. Use of the TEI

In the third section of the survey, we hoped to gain a better understanding of the current uses of the TEI. We first asked participants about the main reason for their interest in the use of the TEI, giving them the following options to be ranked from most to least relevant: 1. TEI Guidelines, 2. Digital Editions, 3. General Text Markup, 4. Articles or other scholarly communications, 5. Metadata Systems, 6. The project where I participated adopted TEI markup, 7. Teaching, 8. Participation in conferences, 9. Other. And the results were as follows:

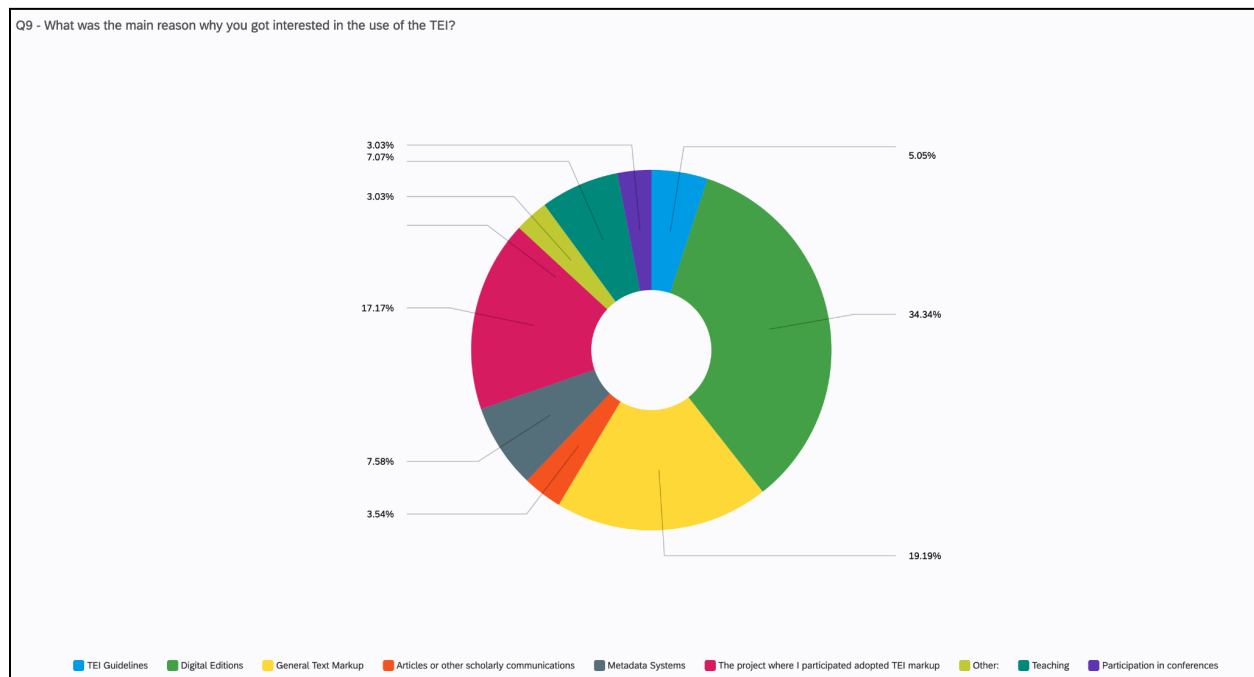


Figure 11. Reasons for being interested in the TEI

As stated in Figure 11, 34.34% of participants (68 responses) chose as the main reason for their interest in the TEI the creation of digital editions. Following, there is their interest in general text markup (19,19%, 38) and—as stated above—their participation in research projects where they are required to work with text encoding (17.1%, 34). Ranked at a lower level, we find interest in Metadata systems (15), teaching (14), in the TEI Guidelines (10) or in articles and other scholarly communications (7). Surprisingly, participation in conferences is rated lower than we might expect (6), since these venues are the best suited for learning about ongoing projects, methods or tools. The remaining 3% chose “Other” and added general terms such as Research, Digital Humanities at large, or Graphs.

The next question, number 10, asked about roles in research projects, and the following options were given: 1. Director / Principal Investigator, 2. Research Team member, 3. Technical Team member, 4. Consulting, 5. Colaborator, 6. I don't participate in any project, 7. N/A. And we obtained the following numbers:

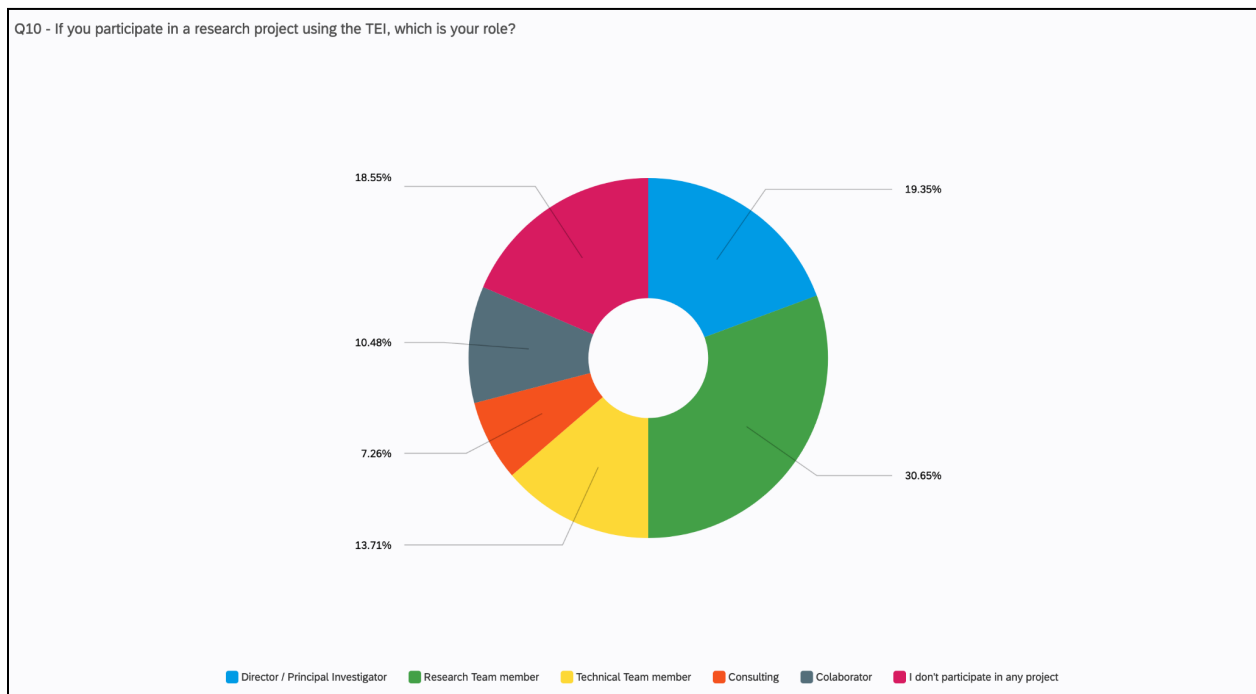


Figure 12. Roles in research projects

The largest part of the participants (30.65%, 38) confirmed that they were members of a Research Team. Directors or Principal Investigators (24) and Technical Team members (17) were among the top answers. Only 13 participants were collaborators in a project using TEI, and 9 offered consulting services. It is curious that a considerable number of participants (18.55%, 23) admitted that they were not participating in a project, which should be taken into consideration as a specific feature of this community and deserves further exploration on the causes that hinder participation.

We gave, at this point, the option of adding the URL of one of the projects where they were participating, and we obtained 46 URLs, which are provided in Appendix 5.

As far as participation in the international TEI community, membership in the TEI Consortium remains extremely low: only 15% of the participants are individual members. The vast majority, 85%, are not members of the TEI. These numbers are very similar to the ones of the 2017 survey. 13% did not answer.

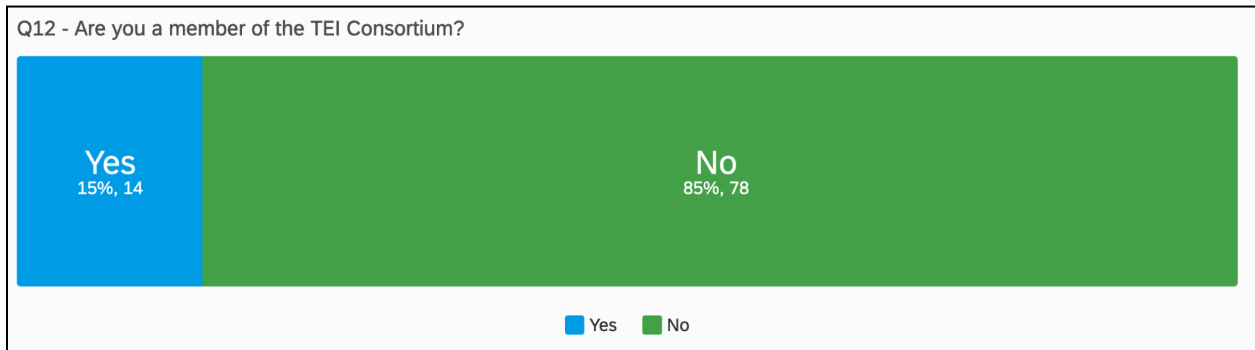
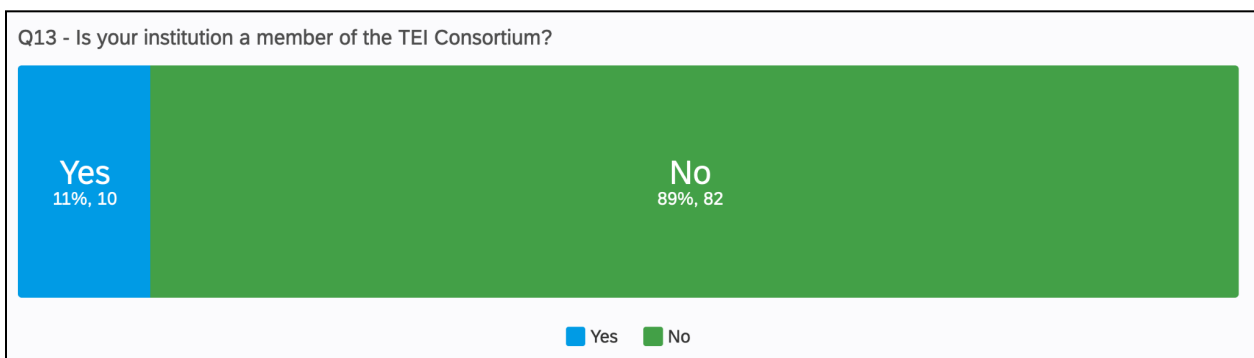


Figure 13. TEI Individual Membership

Additionally, only 9 participants mentioned institutional membership, while 82 are working in institutions that are not members. 13 participants did not answer. The 9 answers related to institutional memberships are from projects outside Latin America.

Figure 14. Institutional membership to the TEI



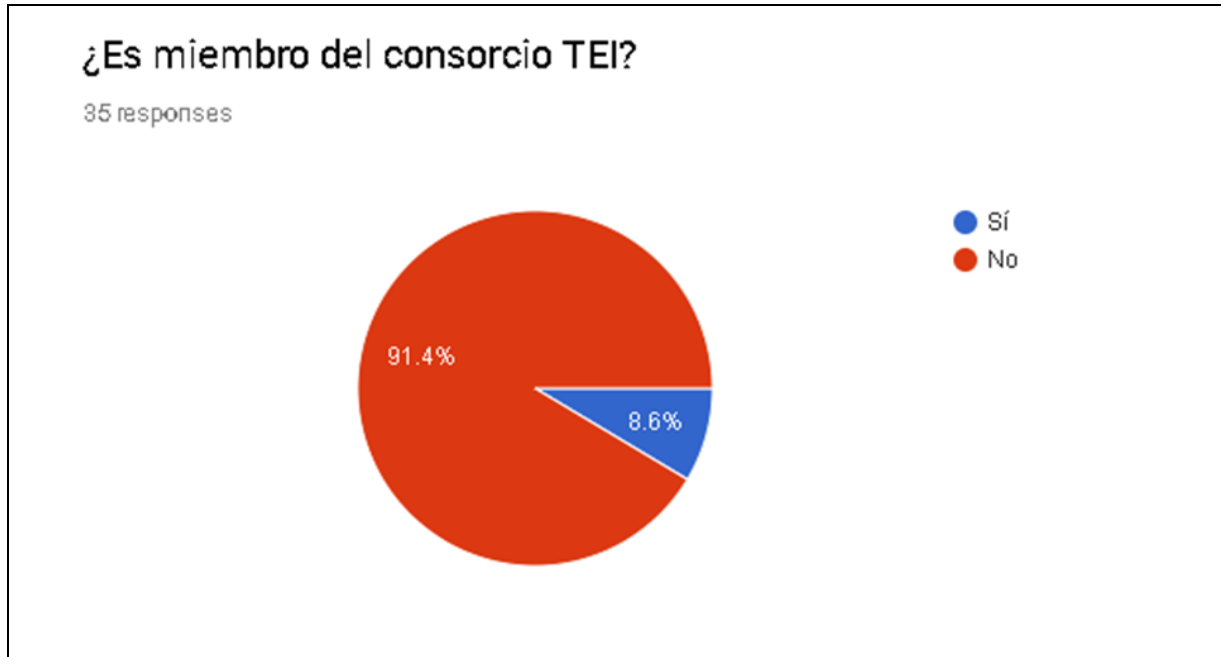


Figure 15. Question on TEI membership. 2017 survey (del Rio Riande 2017)

Participation in TEI conferences seems to have more outreach than membership, though numbers are very low. 21% have participated at least once in a TEI conference; while 70% have never given a paper at the annual TEI conference. 13% did not answer. These numbers are tightly related to the answers on the interests in the TEI in section 2. Overall, we might think that the Hispanophone community remains at the margins of the TEI mainstream, especially in terms of participation as members or in the official events of the Consortium. This situation is probably due not only to the linguistic divide, but also to the lack of infrastructures supporting advanced DH research.

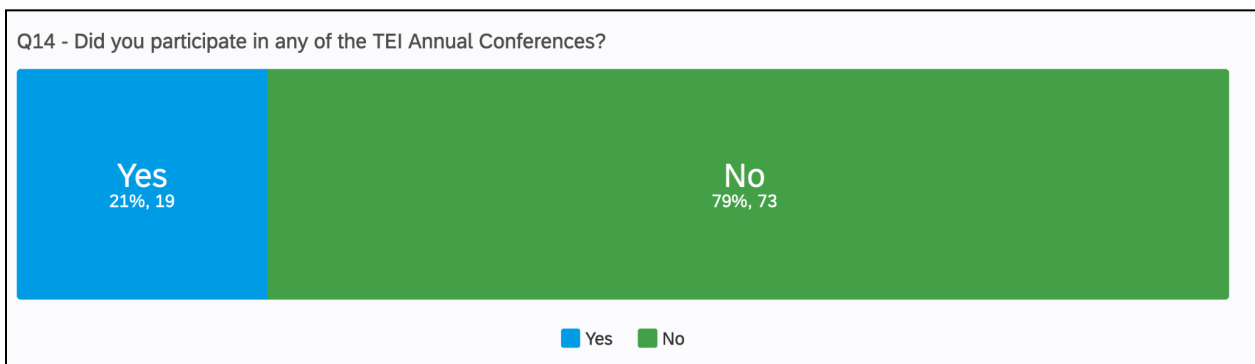


Figure 16. Participation in TEI conferences

Afterwards, we asked participants about the practises used in their projects proposing a multi-choice question that included: 1. TEI customizations, 2. Use of schemas (ODD, RelaxNG, etc.), 3. XML Databases, 4. Relational Databases (MySQL, PostgreSQL, etc.), 5. XSLT Transformations (e.g. from XML to HTML), 6. XQuery Transformations (e.g. from XML to HTML), 7. Controlled vocabularies or thesauri, 8. GIS Technologies, 9. NLP Technologies, 10. Semantic Web Technologies, 11. Data Visualization, or 12. Other. The aim behind some of the choices we gave was to confirm if the TEI Spanish community was in fact aware of best practices already in use at the international level.

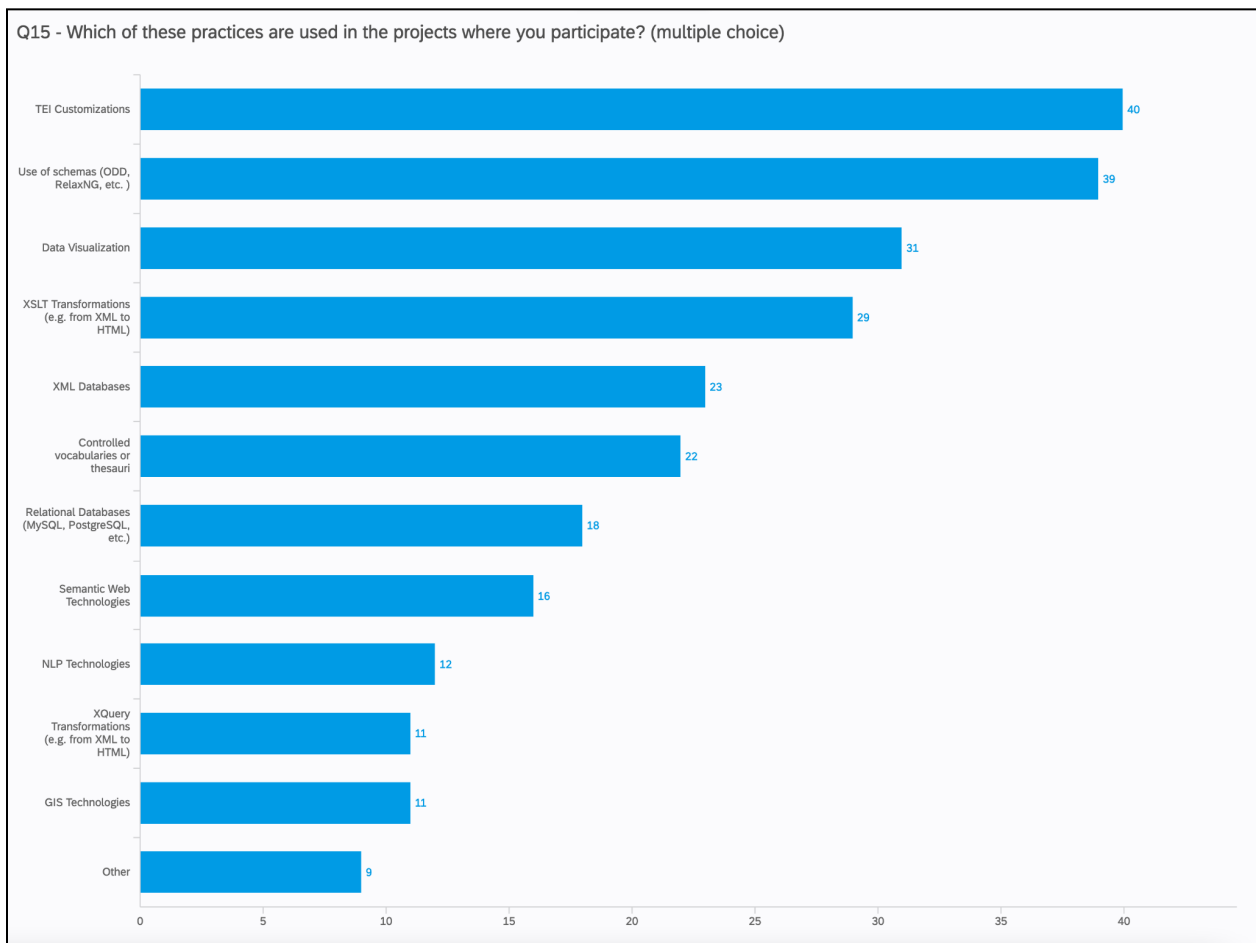


Figure 17. Practices in TEI projects

The answers revealed that most of the projects make use of TEI customizations (40) and they use schemas and ODD (39), which are some of the requirements for a documented, replicable, and sustainable markup workflow.

As far as database systems to organize and recover data, there seems to be a higher use of XML databases (23) ahead of more traditional relational databases (18). This is in line with recent developments, and the arrival of Open Source XML Database products such as eXist.

Concerning XML transformation and rendering, the language most used still seems to be the long-established XSLT (29), while XQuery transformations (11) are less used. This does not seem to be completely in line with the previous question about the use of XML databases, since most native XML databases use XQuery as their primary data retrieval tool, instead of XSLT. There is a curious mix among participants among the “old” and “new” forms.

Other practices which participants adopt when working with the TEI are data visualization (31), and controlled vocabularies or thesauri (22), such as ..., as well as technologies related to the semantic web (16), GIS systems (11), and Natural Language Processing (12%). In “Other,” some participants explained that they were using interoperability scripts between different schemas (DCAT, DDI-CDI), corpora linguistic annotation, XSLT - LaTeX - PDF, and Cocoon. Unfortunately, 9% chose others without further specification.

The following question, number 16, was intended to know if both individual or collective projects shared their XML-TEI files. The reason behind it was to measure participants’ commitment with sharing source code, which we believe helps very much to the advancement and the adoption of established TEI and, in general, DH practices. Although the results are better than we expected, the majority of participants (45) admitted not sharing their files, while 34 answered positively, **and for 26 participants this question did not apply**. In that regard, 29 participants sent the URL of their code (see Appendix 6), and 2 specified that they are still working on it. The fact of not sharing might be due to many factors, among which we can count insecurity in their markup decisions, lack of time to produce a final and optimized version, lack of knowledge of best practices (e.g. choice of public and academic repositories), worries about copyright malpractice, or even copyright restrictions on behalf of their institutions⁶⁴.

⁶⁴ For a list of projects that share their files, see Appendix 6.

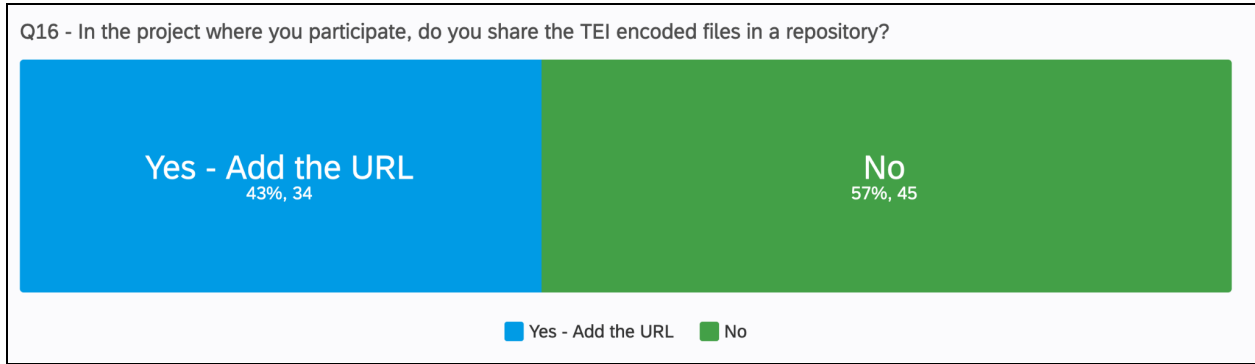


Figure 19. Opening TEI files

The survey also inquired about the use of platforms or tools by participants to publish their TEI files. The goal was to monitor if there was any platform that stands out. Therefore, we proposed the following: 1. Web infrastructure created ad hoc (e.g. XML, XSLT, PHP, Python, etc.), 2. Static Web Generators (Jekyll, Gatsby, etc.), 3. Boilerplate, 4. eXist; 5. TEI Publisher, 6. CETEIcean, 7. Edition Visualization Technology, 8. Other. The answers were as follows:

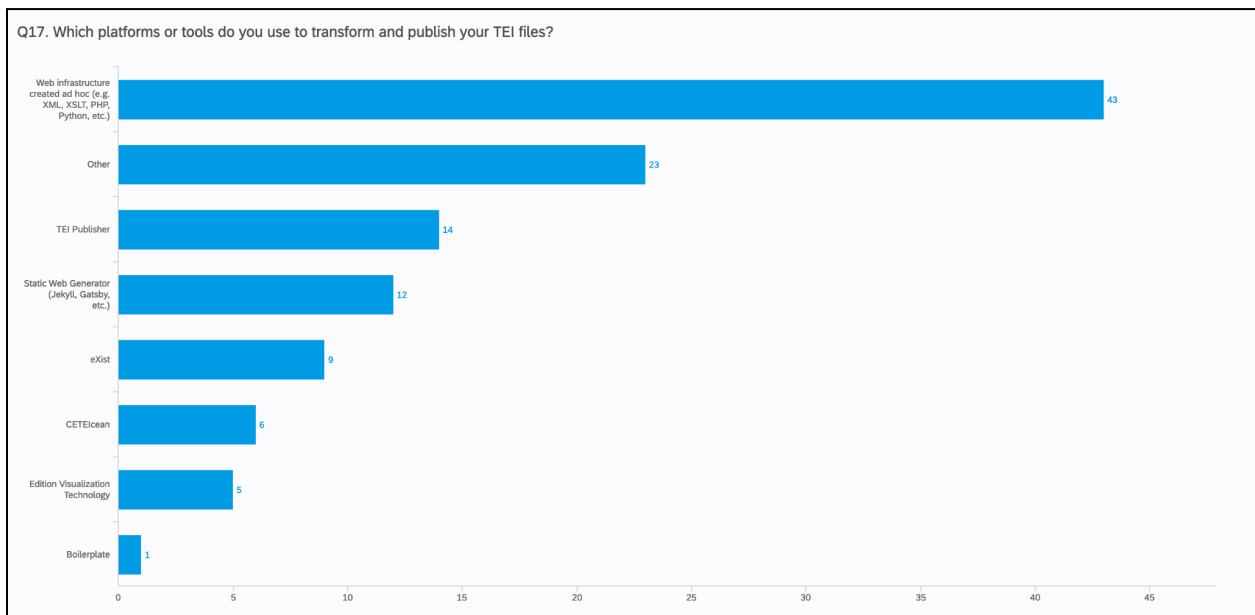


Figure 20. Platforms and tools for transforming and/or publishing TEI

Not surprisingly, the highest score was for infrastructures created *ad hoc* (43). The vast majority hence feel the need to build their own infrastructure in terms of web design, functionalities, search engine, presentation of the text features, and publication. The relevance of this data lies in the dearth of platforms that respond to the needs of practitioners. However,

among the most used platforms, participants choose TEI Publisher⁶⁵ (14), a tool designed back in 2004 to create XML-based repositories through a database with eXist, which is one of the most used native XML databases, and uses a text engine search library called Lucene. In fact, projects relying on the eXist database solution (9) are among the first positions. Following, static web generators are a viable and on the rise option, such as Jekyll or Gatsby (11)⁶⁶. The fact that static web generators are part of the best ranked choices respond probably to two facts: first, the rise of minimal computing that allows the creation of infrastructures without the need of commercial or institutional servers (e.g. static sites can live on free services such as GitHub and GitHub Pages) and, second, the lack of access to digital infrastructure specially in Latin America. Also, among the other platforms to publish the TEI files some participants replied to be using CETEIcean⁶⁷ (6), Edition Visualization Technology⁶⁸ (5), while only one checked the nowadays quite outdated Boilerplate (1). Among those who responded “Other,” they acknowledge their lack of savvy on transforming and publishing TEI files (4), while others mentioned to be using Kiln⁶⁹, TEITOK⁷⁰, Orchid, NetBeans, Cocoon, R scripts, or to have in mind for the future the use of Textual Communities or the above mentioned TEI Publisher.

3.4. Teaching / Learning the TEI

The last block of our survey aimed at gathering some insights about the study and training in text encoding practices, namely the TEI, and to better understand if practitioners were satisfied with the existing resources to teach and learn the TEI.

We started by asking if participants had engaged in teaching the TEI and at what level, offering the following options: 1. Undergraduate, 2. Graduate, 3. Master’s Degree, 4. Summer Schools, 5. Workshops, or 6. I have not taught TEI Courses. These were the answers we obtained:

⁶⁵ TEIPublisher, <http://teipublisher.sourceforge.net/docs/index.php>

⁶⁶ One of these examples could be Ed, <http://elotroalex.github.io/ed/>

⁶⁷ CETEIcean, <https://github.com/TEIC/CETEIcean>

⁶⁸ EVT, <http://evt.labed.unipi.it/>

⁶⁹ Kiln, <https://github.com/kcl-ddh/kiln/>

⁷⁰ TEITOK, <http://www.teitok.org/index.php?action=home>

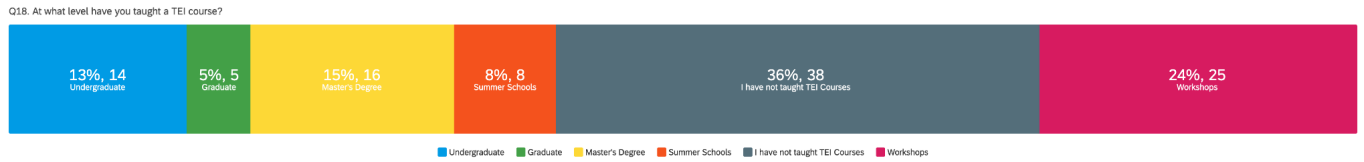


Figure 21. Teaching TEI

The majority of participants (36%, 38) declared not having taught any TEI course. However, among those who had, 24% (25) offered TEI workshops, 15% (16) regular courses at MA level, and 13% (14) undergraduate courses. Less presence seems to have TEI teaching in Summer Schools (8%, 8) and graduate courses (4.72%, 5). As expected, we might conclude by saying that text encoding is not yet part of the established curriculum neither at the undergraduate or graduate academics programs.

In relation to the resources that participants use to teach or to learn the TEI, we gave the following options: 1. Online tutorials and resources; 2. TEI By Example; 3. Text Technologies Hub; 4. DariahTeach; 5. Materials provided by instructors; 6. Scholarly papers; 7. Resources from the TEI-C website, and 8. Other. The numbers were as follows:

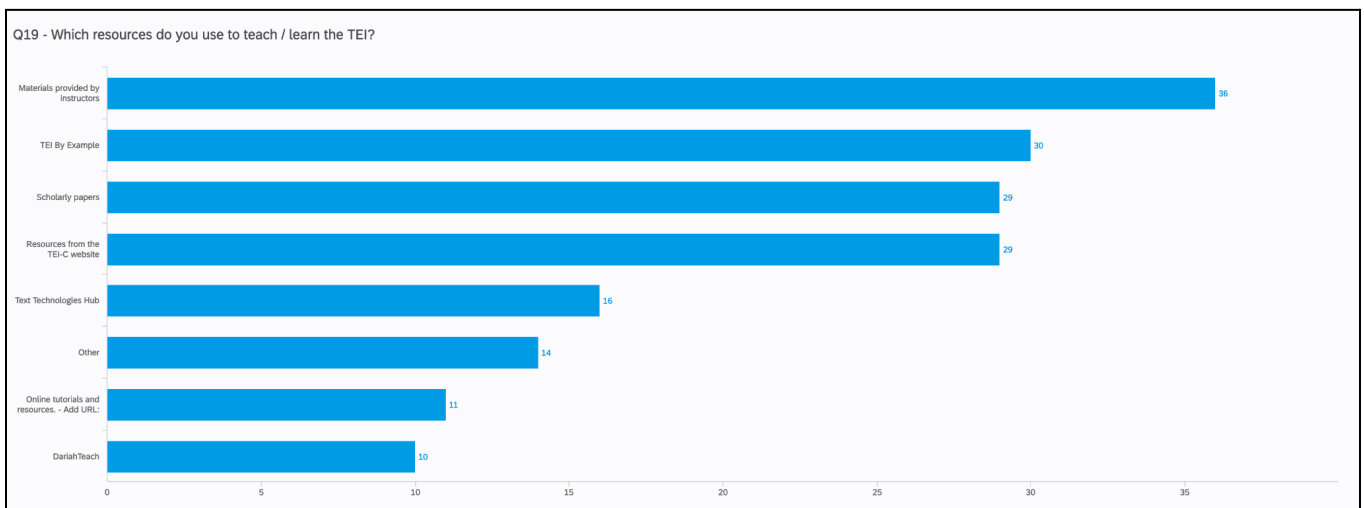


Figure 22. Resources for teaching and learning the TEI

The most common scenario is where instructors provide their own materials (21%, 36), created -we suppose- ad hoc for specific contexts. **This means there is not yet a canonical hub for this type of instruction.** One curious fact that emerges is that the most used resource is *TEI By*

Example (17%, 30), a resource only available in English. This shows that the Spanish speaking world still relies on English resources, followed as well by *DariahTeach* (6%, 10). The only resource in Spanish which offers a not negligible use is our very own *Text Technologies Hub* (9%, 16), that provides materials written in Spanish and conceived for a Hispanophone audience.

Additionally, 29 participants (17%) declared to use the resources listed in the TEI-C website; while the same number turned to scholarly papers dealing with the TEI. Among the answers given under the category “Other” some of the participants insisted on the fact that they use their own material, or that they were not teaching, while two pointed out to the TEI courses at UNED (see above) and one at the University of Birmingham which is connected to an ongoing editing project.

Question 20 inquired about the challenges when working with the TEI. To that end, we gave a set of options that participants needed to rank (1. Lack of infrastructures , 2. Working alone, 3. Lack of a discipline such as Philology or Textual Criticism in the academic curriculum, 4. Lack of funding, 5. Lack of understanding about the TEI in your scholarly community, 6. Impossibility to rely on a hybrid research group (humanists, programmers, librarians, etc), 7. Lack of translation of the TEI Guidelines, 8. Lack of availability of programming and technical training, and 9. Other). And we also gave the possibility to write down additional answers, obtaining the following results:

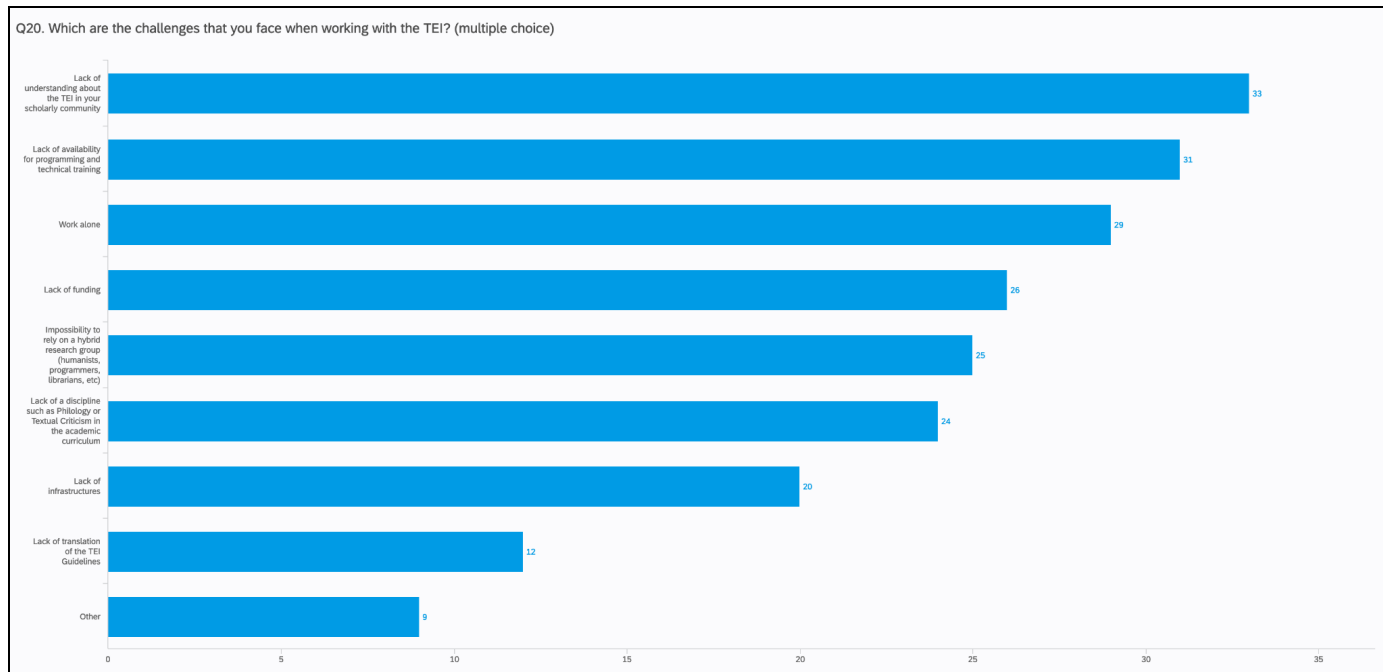


Figure 23. Challenges while working with the TEI

The highest concern among participants seemed to be the fact that in their scholarly community there is still a lack of understanding of what the TEI is (33). Probably connected with this concern, is the uneasiness of individual work (29), with no support from the community, other scholars, or the institution.

The following most ranked challenge faced by participants corresponded to the lack of availability of programming and technical training (31), followed by the lack of funding (26) and the impossibility to rely on a hybrid research group (humanists, programmers, librarians, etc) (25), as well as the lack of infrastructures (20).

Among the options that we offered we wanted the participants to consider the lack of a discipline such as Philology or Textual Criticism in the academic curriculum (24). As we see it, this question is threefold: first, Philology, textual criticism and in general traditional textual scholarship imposed since the late 19th century straightforward methodologies to edit and present the text; this is not yet the case for the digital counterpart, consequently, we lack a clear how-to which is at once set in a theoretical framework; second, in Spain the discipline of textual scholarship has a great tradition, and therefore digital textual practices should had been the

natural development; however, in the academia this adoption of the TEI standard is only beginning; third, this tradition and interest for philology has been scarce or inexistent in Latin America, where most universities do not even include these disciplines in their curricula.

Finally, the lowest ranked, therefore the least important, was the lack of translation of the TEI Guidelines (12). Although this might be an additional barrier to fully adopt this method, altogether considering other data such as the use of the TEI By example for learning purposes, demonstrated that the Hispanophone community does not consider language the main impediment.

In fact, among the answers we received under the option “Other” where respondents who underscored their agreement with all the options given by us (1, 2), others reinforced similar ideas, such as lack of understanding of the field (8) or scarcity of pedagogical materials focused on specific practices (5), while others argued that the challenges come from the youth of the discipline (3), the lack of data in Spanish and clear models (6, 7). However, the most original response was num. 4, in which the participant complains about the approach adopted when working with the TEI: text encoding is not an end in itself, rather a means to produce textual data to be analyzed and processed.

Number	Response	Language
1	All of the above	en
2	Aunque las seleccionaron casi todas, no me lo permite la aplicación	spa
3	El hecho de estar empezando aún.	spa
4	En muchas ocasiones se considera la codificación de textos en TEI como un fin en sí mismo y no como un medio para aportar masivamente textos para diferentes análisis textuales o filológicos	spa
5	La falta de materiales pedagógicos sobre cómo transformar un archivo XML.	spa
6	Lack of data, specially for Spanish	en
7	Lack of recommendations/examples if multiple ways of encoding things are possible; lacking integration of LOD approaches	en
8	Muchos investigadores no asumen la importancia de trabajar con estándares como TEI	spa
9	No lo sé	spa

In Question 21 we asked: “Which kind of resources do you believe we need to improve the teaching and learning of the TEI?” and it was open. The following are the answers we obtained:

Number	Response	Language
1	All of the above	en
2	Bibliografía de referencia clara, dificultad en llegar a niveles avanzados	spa
3	Capacitación	spa
4	Creo que las claves para mejorar la enseñanza y aprendizaje de TEI en la comunidad hispanohablante son la producción de materiales didácticos en español y la formación de equipos de trabajo que integren a especialistas de diferentes áreas, como programación, filología, bibliotecología, etc.	spa
5	Cursos completos (desde planteamiento hasta la salida, pasando por bases de datos y herramientas de visualización web)	spa
6	Cursos y talleres	spa
7	Difficulties of communication between researchers and developers; lack of understanding for the role of a technical editor in the entire research process from the early stage till publication; lack of appreciation for standardization	en
8	Difusión y creación de cursos para personal técnico	spa
9	Económicos y el desarrollo de grupos de trabajo interdisciplinarios	spa
10	Editores de TEI gratuitos que permitan la validación de los documentos en tiempo real	spa
11	Editores XML gratuitos de calidad	spa
12	Editores XML open source y gratuitos	spa
13	Es una pregunta muy general! En general, creo que hacen falta materiales más accesibles de autoaprendizaje. Luego a nivel de cada comunidad hacen falta materiales más personalizados, con explicaciones más amenas a un público general no-técnico.	spa
14	Estrategias de advocacy/difusión más amplias	spa
15	Examples	en
16	Examples	en
17	Financieros	spa
18	Fomentar la transferencia de TEI-XML a TEI-JSON	spa
19	Fomentar y extender su inclusión en el currículo universitario	spa
20	Formación del profesorado proporcionado por las instituciones	spa

21	Formación universitaria	spa
22	Funding!	en
23	Grupos de investigación híbridos	spa
24	Guías en varios idiomas; tutoriales en más idiomas; recursos más allá de las introducciones	spa
25	Materiales pedagógicos sobre XSLT actualizados y en español	spa
26	Mayor difusión con cursos o seminarios	spa
27	Mayor difusión de talleres, cursos para ejercitar la aplicación de TEI y lograr la publicación web. Asistencia técnica	spa
28	Members of the TEI community in academia should work on incorporating TEI in their school curricula	en
29	Money	en
30	No hecho en falta muchos recurso	spa
31	No lo sé	spa
32	Online tutorials	en
33	Open source XML-TEI Editor	en
34	Participación en proyectos con colegas expertos	spa
35	Proyectos de investigación que partan de textos ya codificados en TEI	spa
36	Publicity	en
37	Recursos en español	spa
38	Saber cómo pasar a TEI y qué se puede hacer después con ellos	spa
39	Sobre todo ejemplos, por un lado, y tutoriales para la utilización de los esquemas RNG	spa
40	Software libre, TEI guidelines traducidas, asociación a nivel nacional de TEI.	spa
41	The TEI-C might offer a formal student membership and foster the creation of an online community of students interested in digital editing.	en
42	Traducciones al español de el protocolo TEI	spa
43	Una base centralizada de recursos audiovisuales, no sé si existe	spa
44	Very clear step by step tutorials assuming zero knowledge of computer languages or the terms used to refer to them	en

A closer reading gives us some more clues about the insights of our participants. Some of them have insisted on the need of resources and training materials at all levels and in Spanish,

including as well specific topics (e.g. XSLT) (2, 3, 4, 5, 13, 24, 25, 32, 37, 42, 44), and also providing more examples (15, 16), as well as other types of resources, such as bibliographic references (2). The need for formal and informal training inside and outside academia (6, 8, 20, 21, 26, 27, 28) emerges as well as another legitimate concern. Among the demands for training, it surfaces the lack of knowledge regarding the final stages of the editorial process, that is, the transformation of the TEI files and their online publication (5, 7, 24, 27, 38, 39). This seems to indicate that the community is eager to advance in their technical skills.

It has emerged too the solicitousness for free and reliable XML editors (10, 11, 12, 33, 40), in fact, one of the most popular XML editors in use is proprietary and practitioners, specially from Latin America, cannot afford it or simply do not want to rely on paying software.

Additionally, and as a genuine concern, we find clear demands regarding the need of funding to perform a digital editing project (9, 17, 22, 29) in order to remunerate labor and afford technical support (27).

Finally, some of the participants insisted for the need for outreach at different levels (14, 19, 26, 36, 41) to move forward and, perhaps, to solve the already mentioned lack of understanding from the scholarly community (7), improve dialogue within interdisciplinary teams (7, 9, 23, 34) and even promote real interoperability among other standards (e.g. 18)

Finally, we left the final question, number 22, as open remarks “Is there anything else you would like to add (e.g. field of study, projects, etc. where you use the TEI?” We did not get significant answers. Some of the participants explained their field of research (Medieval, Colonial, Comic studies, Philosophy) and level of studies (grad student working in his dissertation) or role in their current project (digital editor for journals, PI in research projects, consultant, software developer,)

There were, however, two other groups of answers that evince, on one hand, a reticence to adopt the TEI as the standard for encoding within their projects and the lack of confidence on behalf of their team (1, 2); and on the other, the willingness to insist on including TEI training in academic programs (3, 5), to build a stronger community (4), and to do so in the target language (4, 6, 7, 8).

Number	Response	Language
--------	----------	----------

1	Aplico TEI en la realización de bibliotecas digitales: es bueno porque interoperable y difundido; es malo porque impone una estructura jerárquica fija - optamos por la maquetación standoff	spa
2	El proyecto con que colaboro no acaba de convencerse de la necesidad de aplicar TEI de forma profunda y sistemática, posiblemente por falta de formación y tradición.	spa
3	I use the TEI in research and teaching the digital humanities. I would like to see more traditional humanities fields like literature work it into graduate programs as a required unit on scholarly editing methodologies.	en
4	Me parece una buena iniciativa compartir y tener un grupo de TEI en español	spa
5	My project is for my dissertation and I have learned xml in order to transcribe manuscripts. A TEI/ XML course should be a part of every humanities PhD program.	en
6	Sería bueno ver una comunidad de XML-HD en castellano más generalizado, aunque no tengo general [<i>sic</i>] si el XML es tan extendido ahora, o si la manera de uso es igual que hace 10 años (lo uso bastante menos, por ej)	spa
7	Sería interesante congregarse a la comunidad latinoamericana que utiliza TEI, establecer redes y contactos.	spa
8	We need to promote the use of TEI in languages other than English and make available more documentation and examples	en

4. Conclusion and Future work

With a global community with thousands of users, the TEI is one of the most solid ones for text encoding for any type of written source in the Humanities and Social Sciences. The scenarios where TEI is applied vary from digital editing to the creation of linguistic corpus or the creation of XML native databases. The TEI is managed by a Consortium formed by people and institutions, mainly from North America and Europe. This Consortium is responsible for developing and updating the Guidelines for the encoding of texts in digital format. These guidelines are a set of good practices and recommendations that specify general instructions to mark up texts so that they can be processed by computer. The TEI Consortium is insuflated by a spirit of open access and collaboration which means that users themselves are able to propose improvements and modifications through its repository on GitHub. This spirit of collaboration and dissemination is also manifested through a continuous dialogue through a discussion list and the annual organization of the TEI Conference. Finally, the Consortium constituted the umbrella for different working and interest groups for certain topics, such as dictionaries, corpus linguistics, manuscript description, among others.

And yet, the above applies first and foremost to the Anglophone world, where —as we said at the very beginning— the TEI was born. With this report we have tried to delve into a particular community of practitioners, not only defined by their language, Spanish, but also by their cultural and social specific peculiarity and needs. In an international attempt on behalf of the DH community to respect diversity and include languages other than English, it is more than ever clear that languages, as well as cultural knowledge, are necessary if one wishes to understand people and their communities (MLA).

In this report, as introduction, we have provided a succinct state of the art on the Hispanophone DH associations, the main features of digital editions of Spanish texts, and a summary of some of the activities throughout the last years. Although condensed in a few pages, we wanted to bring forward two main facts: first, the relatively long existing tradition of the TEI since the 2000s that brought scholars and institutions to the creation of digital editions and textual corpora. This fact is mainly applicable to Spain where interest for textual scholarship dates back to the early 20th century. And second, that trainings taking place but still in informal contexts (workshop, short courses, summer schools), since TEI is not yet included in formal academic programs.

Our main contribution, also explained succinctly in this report, has been the development of a new platform, the Text Technologies Hub, that fulfills an array of needs that the Hispanophone community has brought forward. That is, the demand for new learning materials, such as online tutorials, repositories of real case scenario examples with TEI markup, access to resources (e.g. tools, catalogs, best practices guidelines, etc.), and the whole while trying to build a Hispanophone community.

However, at the core of this report lies our Survey on the “Use of the Text Encoding Initiative in Spanish for Digital Humanities Projects,” conducted from January until March 2022. Almost a hundred participants have helped us to better understand some insights from different points of view. Geographically speaking, it is now clear to us that most TEI scholars and practitioners are based in Spain, followed by Mexico, the United States, Colombia, and Argentina. Differences in the place of origins and current residence and institution, reflect some global world trends, for example, scholars leave Spain and go elsewhere, USA receives scholars from abroad. The survey has also offered a first scanning of who are those practitioners, mostly academics, but also research support technicians, librarians and archivists, and, at a lesser degree, software developers.

A second big block of questions regarded the scholarly training in the TEI, and it evinced that most of the courses are taken during a Masters and PhD degrees but in informal contexts, since TEI is not part of official programs. Another interesting fact is that practitioners get introduced to the TEI because they are part of a research project.

The third block dealt with the use of the TEI, in other words, why are people using this methodology. Most popular interests are the creation of digital editions, a general interest for markup languages and their digital possibilities, and, again, TEI as technical skills to be hired in a certain project. In terms of engagement, the Hispanophone community does not invest in individual or institutional membership, might this be due to lack of funding or lack of interest remain to be answered. Similar case for the low participation in the annual TEI conference, where almost 80% of the participants have never given a paper. However, some good news emerged from the point of view of digital practices and platforms, which was especially remarkable for the amount of projects sharing their TEI files as open code (Appendix 6).

The third and last block aimed at exploring how people learn and teach the TEI. As for teaching, most practitioners do not teach, and if they do it is in informal contexts, such as workshops, short courses in Masters degree, or summer schools. It has emerged clearly that there are not canonical texts or resources to teach or learn, and most of the participants acknowledge to use Anglophone resources such as *TEI By Example*. In any case, our TTHub seems to be opening the floor to become an unicum within this panorama, since it is the only one with original materiales in Spanish and conceived for it. Finally, we think that we have evince some of the challenges that concern our community, namely the lack of understanding about the TEI among the scholarly community and the lack of availability for programming and digital training. Moreover, we would say that two other concerns have emerged audibly, first, the need for training materials for the whole process of the editorial work, with special insistence on the transformation (XSLT, XQuery) and publication of the TEI files. This means that the community has already overcome the initial stages of familiarity with the TEI, and that now they urge for more advanced topics. Second, several voices have appeared arguing for the need of free XML editors. The fact that the most popular software is proprietary disencourage some users. Needless to say, some steps have already been taken to adapt free softwares to work with TEI files (e.g. a plugin has been created to install in Visual Studio Code).

In sum, the survey allowed us to better understand not only the Spanish-speaking community that uses XML-TEI, but also to think of strategies that can contribute with more inclusive practices for scholars from less represented countries and in less favorable contexts inside the global TEI community. Last but not least, we believe the survey will be useful for designing actions that can support a wider range of modes of interaction and collaboration inside the global TEI community.

The Survey and the development of the TTHub have been accompanied by multiple activities that we have organized under the auspices of this Mellon grant:⁷¹ participation in international conferences (a Roundtable at the V Conference of the HDH, in Santiago de Compostela, Spain (Online), a Roundtable and a Short paper at the TEI Conference and Conference's Meeting in 2021 (Online) and a Long paper at the same TEI conference in 2022 in Newcastle, a workshop and a Long paper at the V Conference of the AAHD, General Roca, Argentina), several workshops (at the CSIC in Madrid, and at the Universidad Católica de

⁷¹ See Appendix 7 for the full list.

Valencia), outreach activities (Redes TEI (Tw/Fb), a Editaton at the Semana HD). Additionally, we have also submitted an article under the title “Autonomía y control: Minimal Computing como propuesta pedagógica para las Humanidades Digitales” for a companion of pedagogy materials to teach Hispanic language and culture.

Finally, we are currently preparing a special issue in Spanish at the *Journal of the Text Encoding Initiative* with eight contributions already submitted and ten authors.

In the months to come, we wish to apply for additional funding to bring forward the development of the translation tool, the optimization of other features in the TTHub (e.g. an interactive workspace, and the delivery of our scripts in open source), as well as keep organizing scholarly and outreach events.

5. Bibliography

Allés-Torrent 2015, 2017, 2020;
Allés-Torrent 2017
del Rio Riande 2015,
del Rio Riande 2015a,
del Rio Riande 2015b,
del Rio Riande 2016a
del Rio Riande 2016a,
del Rio Riande 2016b
del Rio Riande and Zubillaga
2015;
del Rio Riande, 2015, 2016ab
Faulhaber 1996;
Fiormonte 2014
Fiormonte and del Rio Riande
2017
Fradejas 2009;
Galina 2013
González-Blanco et al. 2014,
2015;
González-Blanco et al. 2017a
Isasi & Rojas Castro 2021
Lucía Megías 2008
Marcos Marín 1986,
Marcos Marín 1986;
Marcos Marín 1994
O'Donnell 2012
O'Donnell et al. 2016a
O'Donnell et al. 2016b
Priani 2017

Revenga 2014;

Rojas Castro 2017

Spence & Brandao 2020

Spence 2014;

Spence and González-Blanco 2014;

TEI Consortium, P5: Guidelines for Electronic Text Encoding

and Interchange. Version 4.4.0 (Last update April 19th, 2022).

<https://tei-c.org/release/doc/tei-p5-doc/en/html/index.html>

Terras, Van Den Branden, and Vanhoutte 2009: 299

6. Appendices

Appendix 1: Survey Questions

[English Version]

Survey Title: Use of the Text Encoding Initiative in Spanish for Digital Humanities Projects

Introduction:

This survey is part of the Mellon funded project “Communicating the Text Encoding Initiative to a Multilingual User Community” and aims to better understand the needs of the Spanish-speaking TEI community.

The survey consists of 22 questions, and should not take more than 10 minutes to complete. The Survey is anonymous, although we ask about your affiliation and nationality in order to obtain a better understanding from a geographical perspective. The Survey’s data will be used for a Report on the uses of the TEI and the needs of the Spanish-speaking community.

If you have used the TEI for your research or taken any TEI course at any level, and you are either part of the Global Spanish-speaking community or work with Spanish texts encoded in TEI, you are kindly invited to answer these questions.

Survey Questions:

Would you like to take this survey and give consent for your answers to be used for academic purposes?

- Yes
- No

I. General Information

Institution to which you currently belong

Which city and country do you currently live in (e.g. Buenos Aires, Argentina)?

What is your country of origin?

What is your current position?

- Assistant Professor
- Associate Professor
- Full Professor
- Lecturer
- Librarian
- Researcher
- MA or Graduate Student
- Postdoc
- Developer
- Other:

II. Scholarly Training and TEI

I studied or worked with TEI at the level of...

- PhD Degree
- Master's Degree
- Undergraduate studies (I took a TEI course)
- Library offering support for TEI projects
- Other

Have you taken MA or Graduate courses related to the TEI?

- Yes
- No

[If YES]

In which country did or do you study your MA or graduate TEI training?

Did your MA or PhD advisor belong to your university?

- Yes
- No
- I had a co-advisor who directed the TEI work

III. Use of the TEI

What was the main reason why you got interested in the use of the TEI? (multiple choice)

- TEI Guidelines
- Digital Editions
- General Text Markup
- Articles or other scholarly communications
- Metadata Systems
- The project where I participated adopted TEI markup
- Teaching
- Participation in conferences
- Other:

If you participate in a research project using the TEI, which is your role? (multiple choice)

- Director / Principal Investigator
- Research Team member
- Technical Team member
- Consulting
- Colaborator
- I don't participate in any project

URL of one of the projects

--

Are you a member of the TEI Consortium?

- Yes
- No

Is your institution a member of the TEI Consortium?

- Yes
- No

Did you participate in any of the TEI Annual Conferences?

- Yes
- No

Which of these practices are used in the projects where you participate? (multiple choice)

- TEI Customizations
- Use of schemas (ODD, RelaxNG, etc.)
- XML Databases
- Relational Databases (MySQL, PostgreSQL, etc.)
- XSLT Transformations (e.g. from XML to HTML)
- XQuery Transformations (e.g. from XML to HTML)
- Controlled vocabularies or thesauri
- GIS Technologies
- NLP Technologies
- Semantic Web Technologies
- Data Visualization
- Other

In the project where you participate, do you share the TEI encoded files in a repository?

- Yes - Add the URL
- No

Which platforms or tools do you use to transform and publish your TEI files? (multiple choice)

- Web infrastructure created ad hoc (e.g. XML, XSLT, PHP, Python, etc.)
- Static Web Generator (Jekyll, Gatsby, etc.)

- Boilerplate
- eXist
- TEI Publisher
- CETEIcean
- Edition Visualization Technology
- Other

IV. Teaching / Learning the TEI

At what level have you taught a TEI course?

- Undergraduate
- Graduate
- Master's Degree
- Summer Schools
- Workshops
- I have not taught TEI Courses

Which resources do you use to teach / learn the TEI?

- Online tutorials and resources. - Add URL:
- TEI By Example
- Text Technologies Hub
- DariahTeach
- Materials provided by instructors
- Scholarly papers
- Resources from the TEI-C website
- Other

Which are the challenges that you face when working with the TEI? (multiple choice)

- Lack of infrastructures
- Work alone
- Lack of a discipline such as Philology or Textual Criticism in the academic curriculum
- Lack of funding
- Lack of understanding about the TEI in your scholarly community
- Impossibility to rely on a hybrid research group (humanists, programmers, librarians, etc)
- Lack of translation of the TEI Guidelines
- Lack of availability for programming and technical training
- Other

Which kind of resources do you believe we need to improve the teaching and learning of the TEI?

Is there anything else you would like to add (e.g. field of study, projects, etc. where you use the TEI)?

Gracias por dedicarle tiempo a esta encuesta. /Thank you for taking the time to complete this survey.

Su respuesta se ha registrado. / Your answer has been recorded.

[Spanish Version]

Survey Title: Uso de la Text Encoding Initiative en español para proyectos de Humanidades Digitales

Introduction:

Esta encuesta forma parte del proyecto “Communicating the Text Encoding Initiative to a Multilingual User Community” financiado por la Fundación A. Mellon y tiene como objetivo comprender mejor las necesidades de la comunidad TEI hispanohablante.

Este cuestionario consta de 22 preguntas, por lo que no debería llevar más de 10 minutos completarla. Se trata de una encuesta anónima, si bien solicitamos datos de su filiación y nacionalidad para obtener una mejor comprensión desde el punto de vista geográfico. Los datos de la encuesta se utilizarán para elaborar un informe sobre los usos de la TEI y las necesidades de la comunidad de habla hispana.

Le invitamos a responder estas preguntas especialmente si ha utilizado TEI para sus investigaciones o ha tomado algún curso sobre TEI a cualquier nivel, o si forma parte de la comunidad TEI hispanohablante global o trabaja con textos en español codificados con TEI.

Survey Questions:

¿Desea realizar esta encuesta y dar consentimiento para que sus respuestas sean utilizadas para fines académicos?

- Si
- No

I. Inforamción general

Institución a la que pertenece en la actualidad

¿En qué ciudad y país reside en la actualidad (ej. Buenos Aires, Argentina)?

¿Cuál es su país de nacimiento?

¿Cuál es su cargo actual?

- Profesor Ayudante / Asistente
- Profesor Titular
- Profesor Catedrático
- Profesor Contratado
- Bibliotecario
- Investigador
- Estudiante de Máster o Doctorado
- Becario Postdoctoral
- Desarrollador
- Otro:

II. Formación académica y TEI

He realizado estudios o trabajado con TEI a nivel de...

- Una tesis doctoral
- Una tesis de maestría
- Soy estudiante de grado y tomo un curso de TEI
- Soy bibliotecario y doy apoyo a proyectos sobre TEI
- Otro

¿Ha realizado estudios de maestría o doctorado relacionados con TEI?

- Si
- No

[If YES]

¿En qué país realiza o realizó sus estudios de maestría o doctorado?

¿Su director de tesis (de doctorado o maestría) pertenecía a su universidad?

- Sí

- No
- Tenía un co-director de otra universidad que dirigía lo relacionado con la TEI

III. Uso de la TEI

¿Cuál es el motivo principal por el que se interesó en el uso de la TEI? (opción múltiple)

- Las Guías directrices (Guidelines)
- La edición digital
- La codificación de textos en general
- Publicaciones periódicas
- Sistemas de metadatos
- El proyecto en el que colaboraba utilizaba codificación con TEI
- Enseñanza
- Participación en conferencias
- Otro

Participa en algún proyecto de investigación que utilice TEI en calidad de ... (opción múltiple)

- Director
- Miembro del equipo investigador
- Miembro de equipo técnico
- Consultor
- Colaborador externo
- No participo en ningún proyecto

URL de alguno de los proyectos:

¿Es usted miembro del Consorcio TEI?

- Sí
- No

¿Es la institución a la que pertenece miembro del Consorcio TEI?

- Sí
- No

¿Participó en alguno de los congresos anuales de la TEI?

- Sí
- No

En los proyectos que dirige o en los que participa se utilizan (opción múltiple)

- Personalización de los módulos TEI
- Uso de esquemas (ODD, RelaxNG, etc.)
- Bases de datos XML nativas
- Bases de datos relacionales (MySQL, PostgreSQL, etc.)
- Transformaciones en XSLT
- Transformaciones en XQuery
- Vocabularios controlados o tesauros
- Tecnologías relacionadas con GIS
- Tecnologías relacionadas con NLP
- Tecnologías relacionadas con Web semántica
- Visualización de datos
- Otro:

En el proyecto que participa, ¿se comparten en un repositorio los archivos codificados en TEI?

- Sí - Agregue el URL
- No

¿Qué plataformas o herramientas utiliza para la transformación y publicación de sus archivos TEI? (opción múltiple)

- Entorno web creado ad hoc (e.g. XML, XSLT, PHP, Python, etc.)
- Generadores de web estáticas (Jekyll, Gatsby, etc.)
- Boilerplate

- eXist
- TEI Publisher
- CETEIcean
- Edition Visualization Technology
- Otros

IV. Enseñanza / Aprendizaje de TEI

Ha enseñado cursos sobre TEI a nivel de...

- Estudios de grado
- Doctorado
- Máster
- Cursos de verano
- Talleres
- No he enseñado cursos sobre TEI

¿Qué recursos utiliza para enseñar / aprender TEI?

- Tutoriales y materiales online. - Agregue la URL:
- TEI By Example
- Text Technologies Hub
- DariahTeach
- Materiales provistos por profesores
- Artículos de investigación
- Materiales del sitio de la TEI-C
- Otros

¿Cuáles son los obstáculos a los que se enfrenta a la hora de trabajar con TEI? (opción múltiple)

- La escasez de infraestructuras
- El trabajo aislado
- La ausencia de una disciplina como la Filología o la Crítica Textual en el curriculum universitario
- La falta de financiación

- La falta de conocimiento sobre el tema en su comunidad científica
- La imposibilidad de contar con un grupo de investigación híbrido (humanistas, programadores, bibliotecarios, etc)
- La falta de traducciones de materiales como las Guidelines de la TEI
- La necesidad de capacitación propia en temas de programación
- Otros

¿Qué tipo de recursos cree que faltan para mejorar la enseñanza y el aprendizaje de TEI?

¿Quiere hacernos algún otro comentario (ej. campo de estudio donde aplica TEI, proyectos en los que usa TEI, etc.)?

Gracias por dedicarle tiempo a esta encuesta. /Thank you for taking the time to complete this survey.

Su respuesta se ha registrado. / Your answer has been recorded.

Appendix 2: Invitation email

[English Version]

Survey: Use of the TEI in Spanish

Dear list members,

As part of the Mellon-funded initiative “Communicating the TEI to a multilingual community,” we are researching the contexts and uses of the Text Encoding Initiative among the global Spanish-speaking community. If you have used the TEI for your research or taken any TEI course at any level, and you are either part of the Global Spanish-speaking community or work with Spanish texts encoded in TEI, you are kindly invited to answer these questions.

We are interested in surveying the different scenarios where TEI is used, the geographical diversity, and needs for training and learning resources. We are launching this survey in the hopes of covering all those Spanish-speaking areas, but also all those projects and users from other regions working with Spanish primary sources (e.g. projects in the US, using primary sources in Spanish, etc.).

The survey consists of 22 questions and should not take more than 10 minutes to complete. The survey is anonymous, although we ask about your affiliation and nationality in order to obtain a better understanding from a geographical point of view (we do not ask for demographic information such as age, gender, ethnicity or religion).

The survey’s data will be used for a Report on the uses of the TEI and the needs of the Spanish community.

The survey can be answered either in Spanish or English and will remain open until February 28, 2022. Please share it with colleagues and friends who might be able to contribute!

Follow this link to the Survey: <https://bit.ly/encuestaTEI> or copy and paste the URL into your internet browser: https://umiami.qualtrics.com/jfe/form/SV_aWx84qH6cih9Xf0

We really appreciate your collaboration and we truly value the information you can provide us.

Do you have questions? Please email our research team via contacto@tthub.io or directly to Susanna Allés-Torrent <susanna_alles@miami.edu> or Gimena del Rio Riande <gdelrio@conicet.gov.ar>

Thank you very much,

Susanna Allés Torrent & Gimena del Rio Riande

[Spanish Version]

Encuesta: Uso de la TEI en español

Estimados miembros de la lista,

Como parte de la iniciativa financiada por la Fundación Andrew W. Mellon "Communicating the Text Encoding Initiative to a Multilingual User Community", estamos investigando los contextos y usos de la Text Encoding Initiative en la comunidad hispanohablante. Si ha utilizado TEI para su investigación o ha realizado algún curso sobre TEI de cualquier nivel, y/o forma parte de la comunidad TEI hispanohablante global o trabaja con textos en español codificados en TEI, le invitamos a responder a estas preguntas.

Nos interesa conocer los diferentes escenarios en los que se utiliza TEI, la diversidad geográfica, las necesidades de formación así como los recursos de aprendizaje. Lanzamos esta encuesta con la esperanza de abarcar todas aquellas zonas de habla hispana, pero también todos aquellos proyectos y usuarios de otras regiones que trabajen con fuentes primarias en español (por ejemplo, proyectos en EE.UU. que utilicen fuentes primarias en español, etc.).

Este cuestionario consta de 22 preguntas, por lo que no debería llevar más de 10 minutos completarlo. Se trata de una encuesta anónima, si bien solicitamos datos de su filiación y nacionalidad para tener una mejor comprensión desde el punto de vista geográfico (no pedimos información demográfica como edad, sexo, etnia o religión).

Los datos de la encuesta se utilizarán para elaborar un informe sobre las necesidades de la comunidad de habla hispana sobre el uso de la TEI.

La encuesta puede contestarse en español o inglés seleccionando la lengua de preferencia, y permanecerá abierta hasta el 28 de febrero de 2022. Por favor, ¡compártala con colegas y amigos que puedan contribuir!

Siga este enlace a la encuesta: <https://bit.ly/encuestaTEI>, o copie y pegue la URL en su navegador de Internet: https://umiami.qualtrics.com/jfe/form/SV_aWx84qH6cjh9Xf0

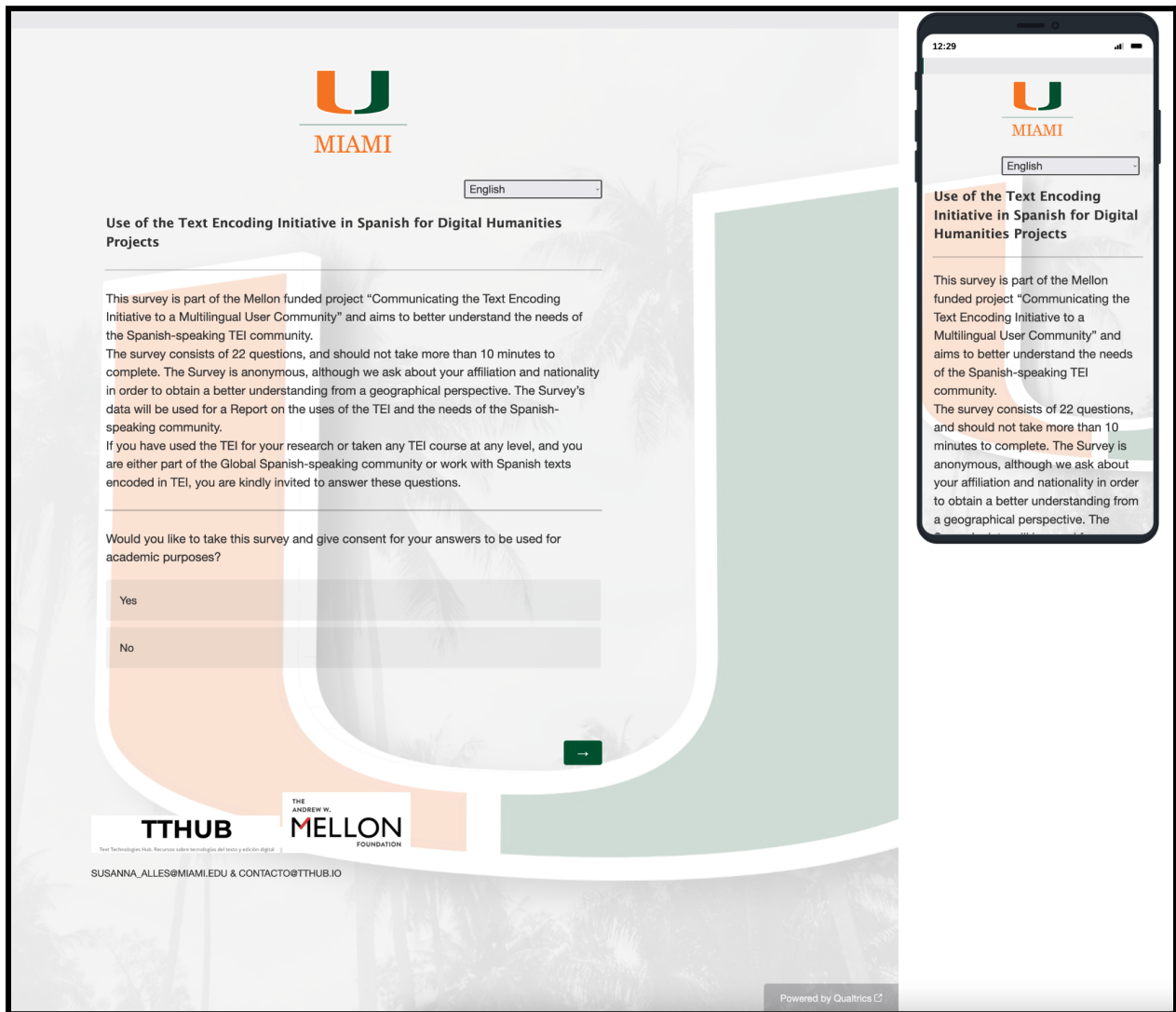
Apreciamos su colaboración y valoramos mucho la información que nos pueda proporcionar.

Si tiene alguna pregunta, envíe un correo electrónico a nuestro equipo de investigación a través de contacto@tthub.io o directamente a Susanna Allés-Torrent <susanna_alles@miami.edu> o Gimena del Rio Riande <gdelrio@conicet.gov.ar> .

Muchas gracias

Susanna Allés Torrent & Gimena del Rio Riande

Appendix 3: Screenshot of the Survey



Appendix 4: Figures

Figure 1. Countries of Residence of Participants

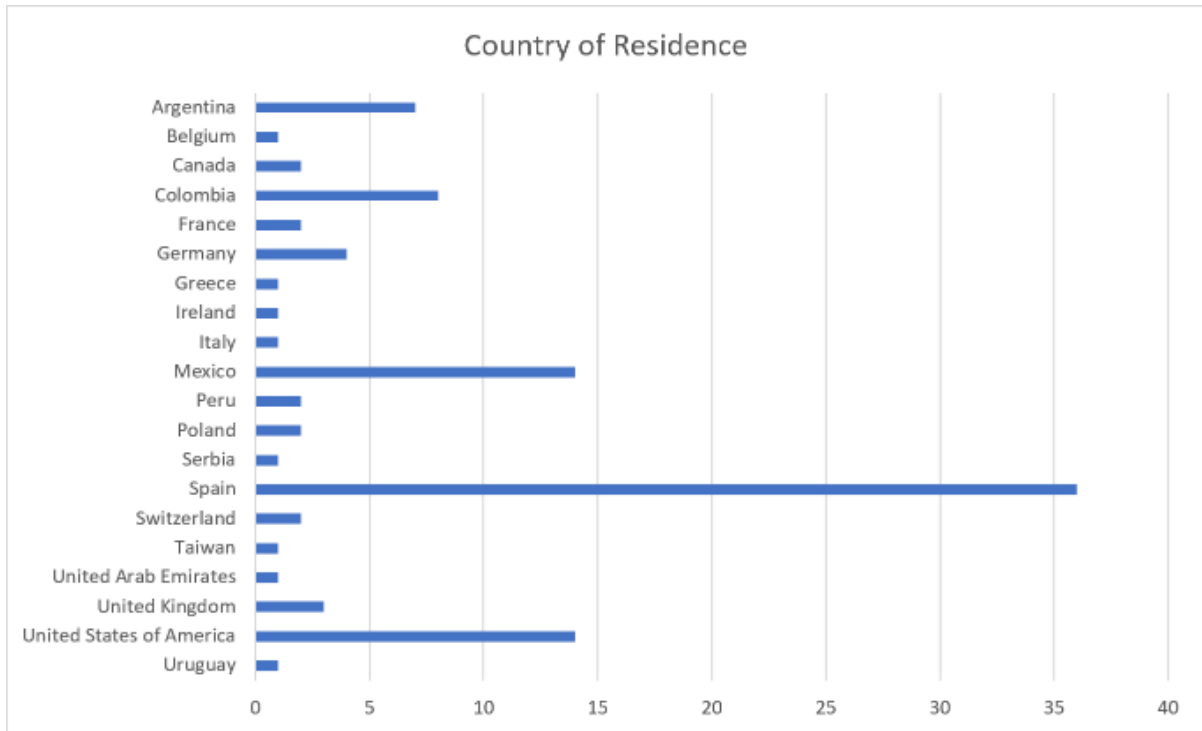


Figure 2. Countries of Origin of Participants

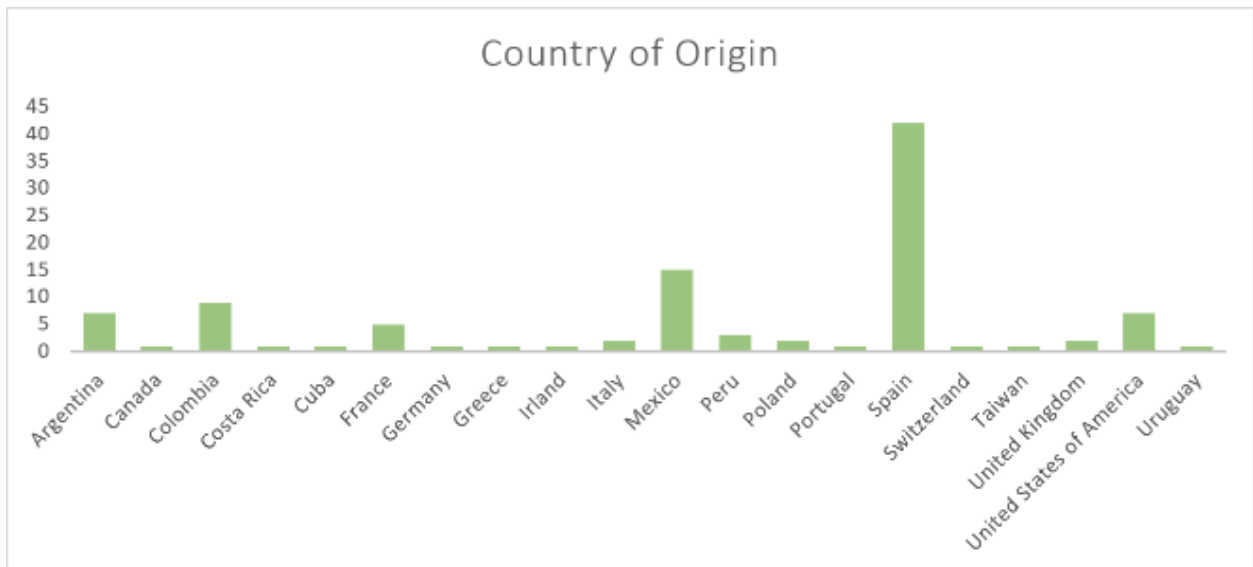


Figure 3. Participation by Language

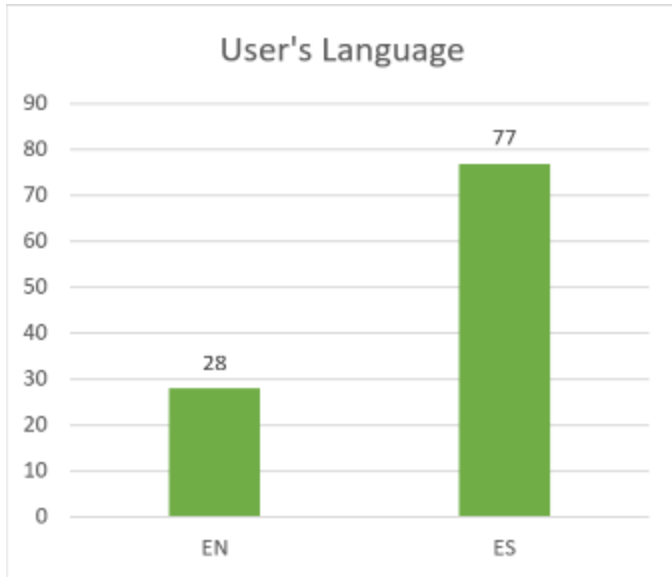


Figura 4. Academic Positions of the Participants

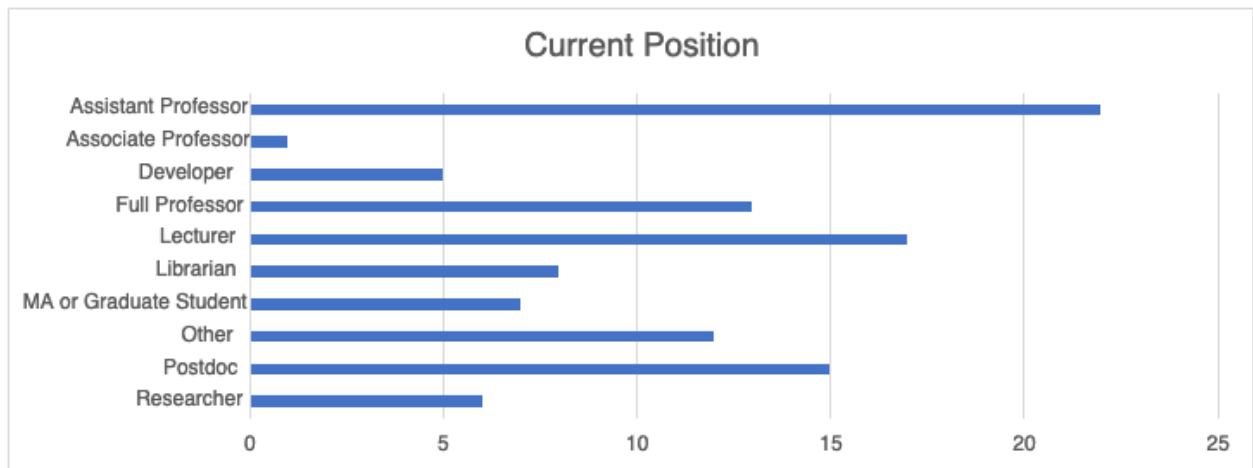


Figura 5. Scholarly Training of Participants

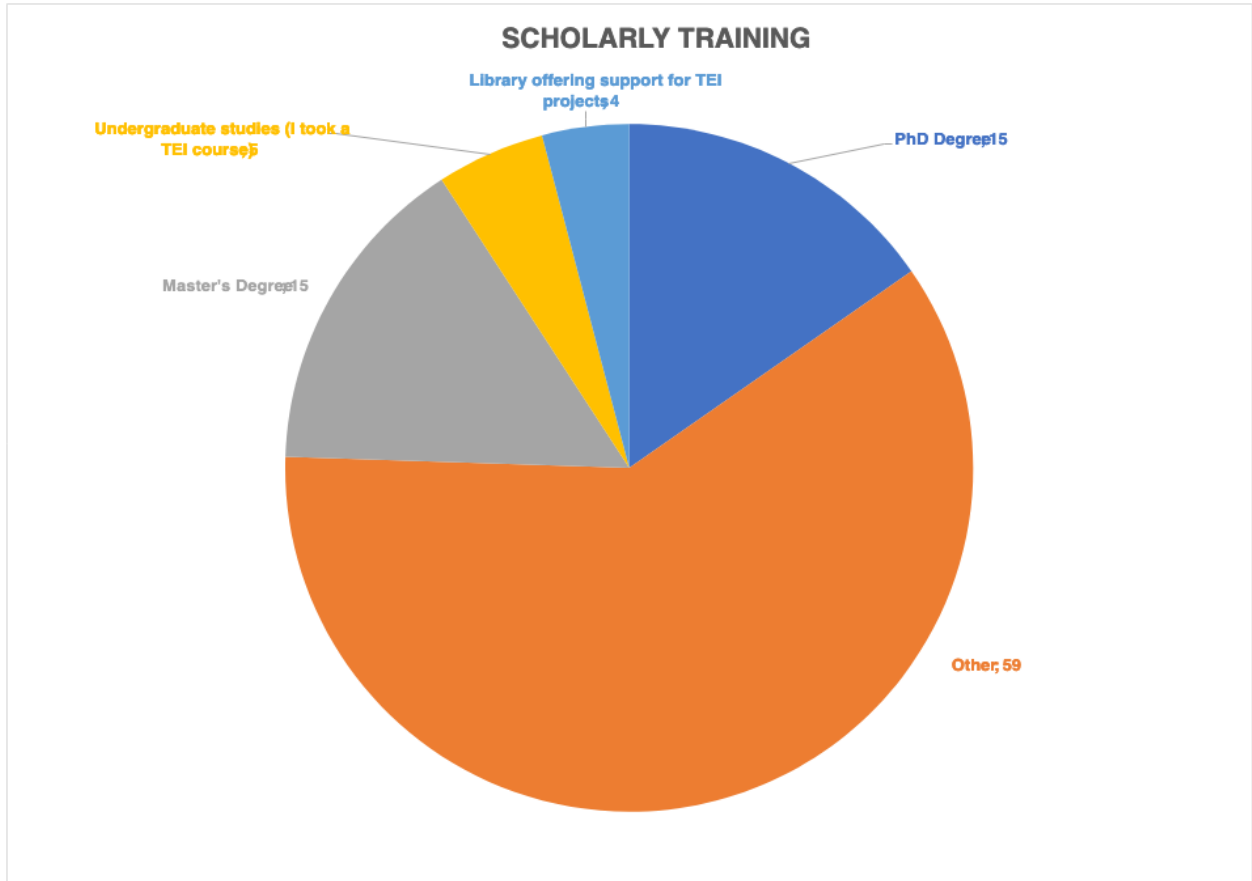


Figure 6. MA or Graduate Courses taken by the Participants

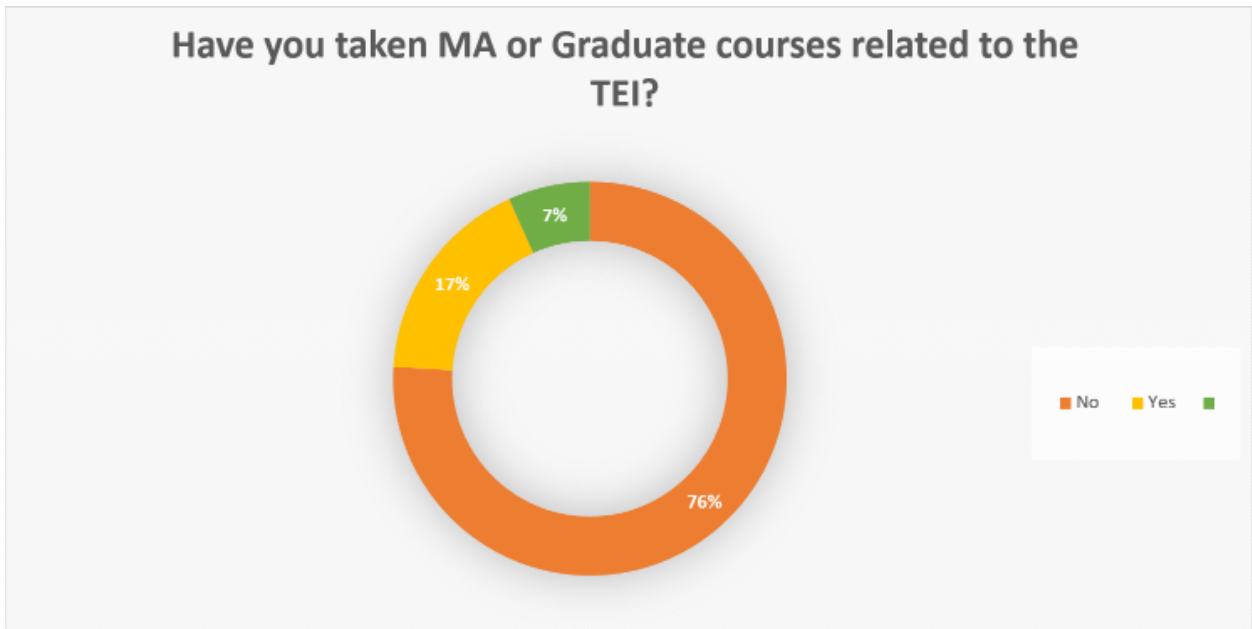


Figure 7. Country where participants studied the TEI

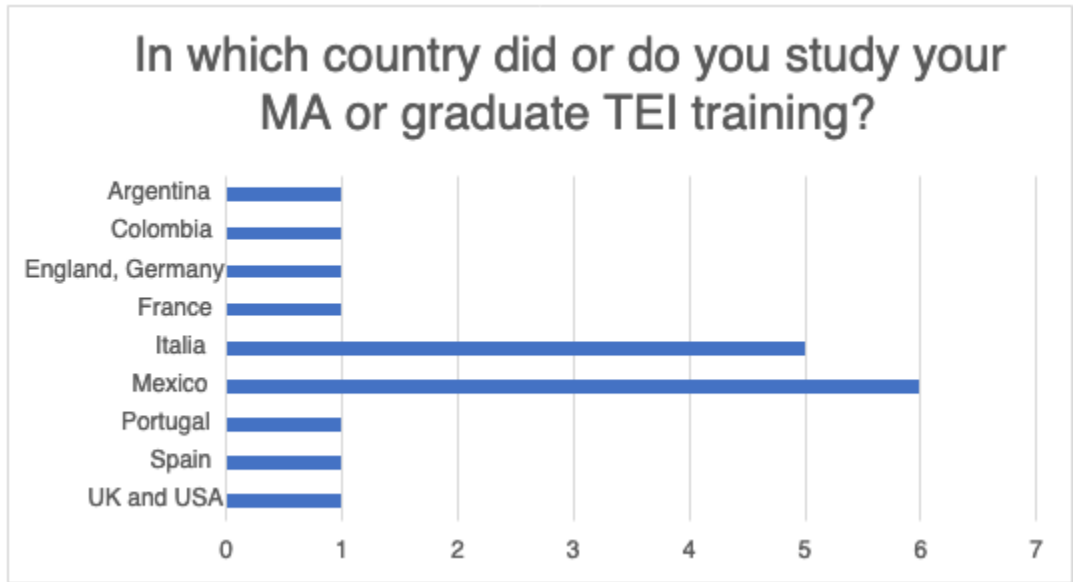


Figure 8. MA or PhD advisor belonging to the same institution

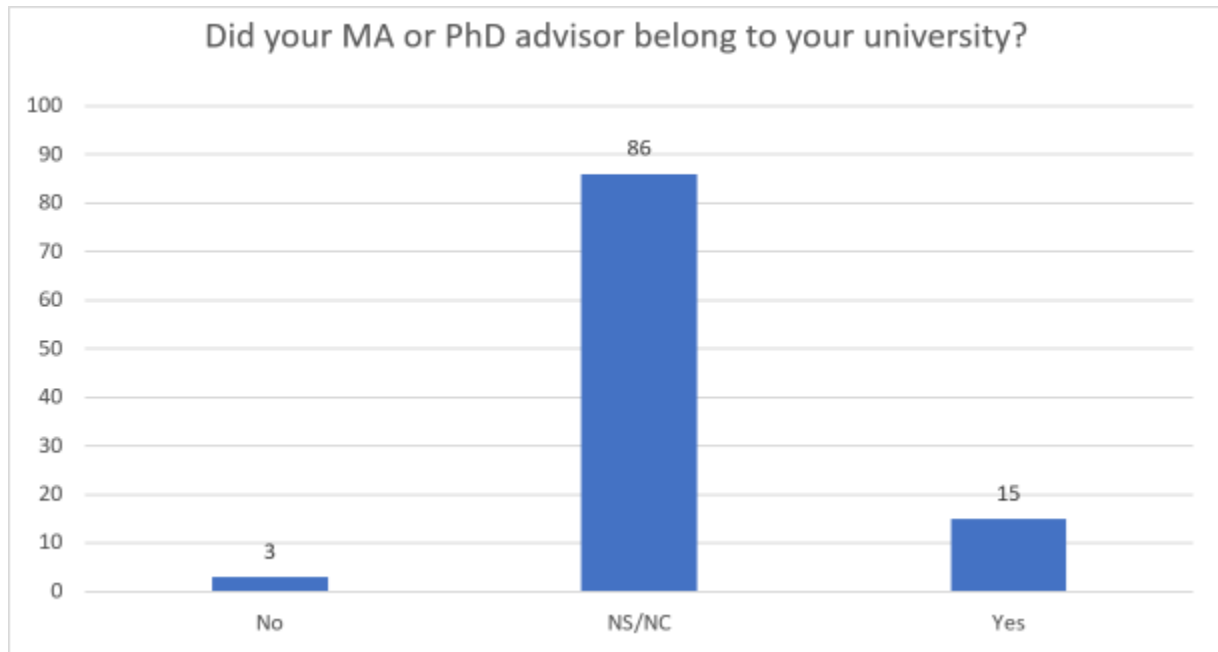


Figure 9. Role in Research Projects

ROLE IN A RESEARCH PROJECT

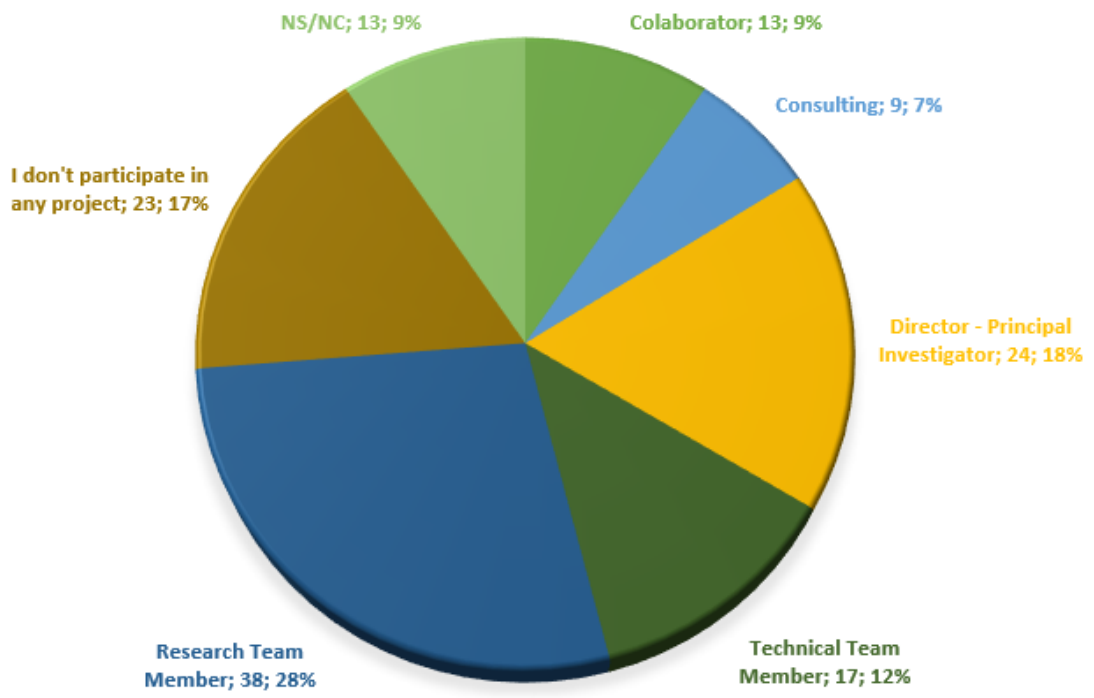


Figure 10. Reason of interest in Using the TEI

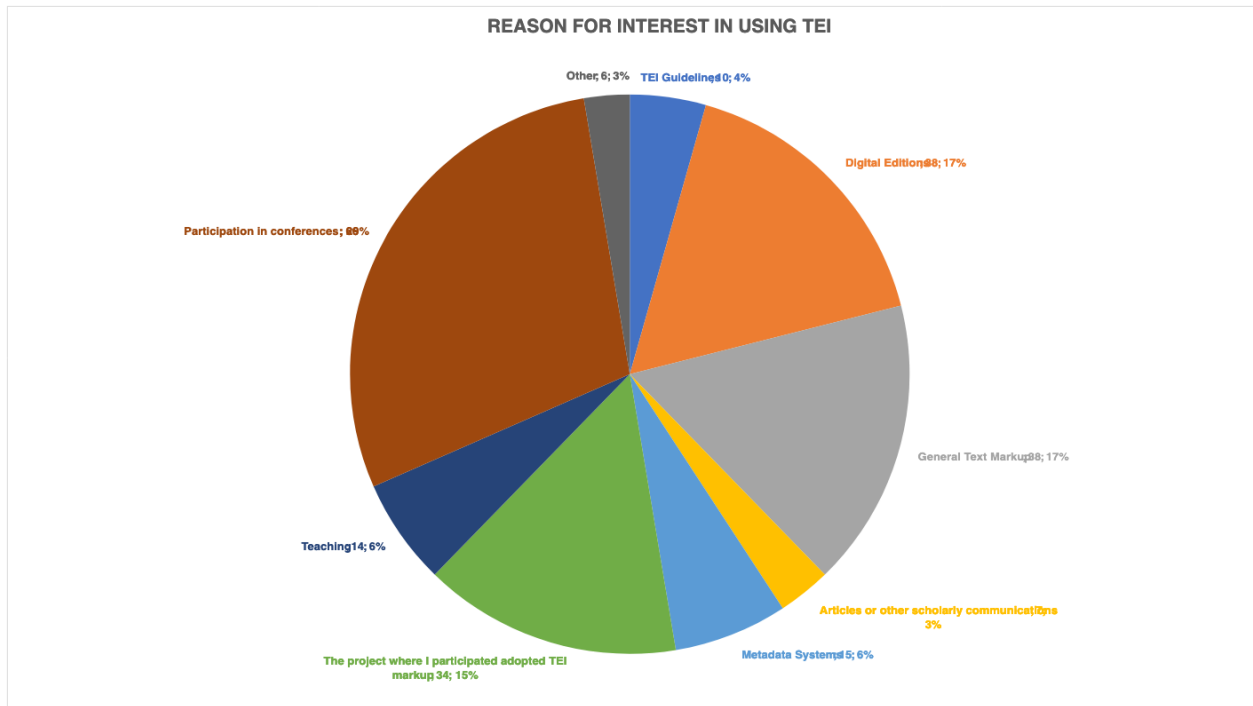


Figure 11. Consortium Membership: Individuals

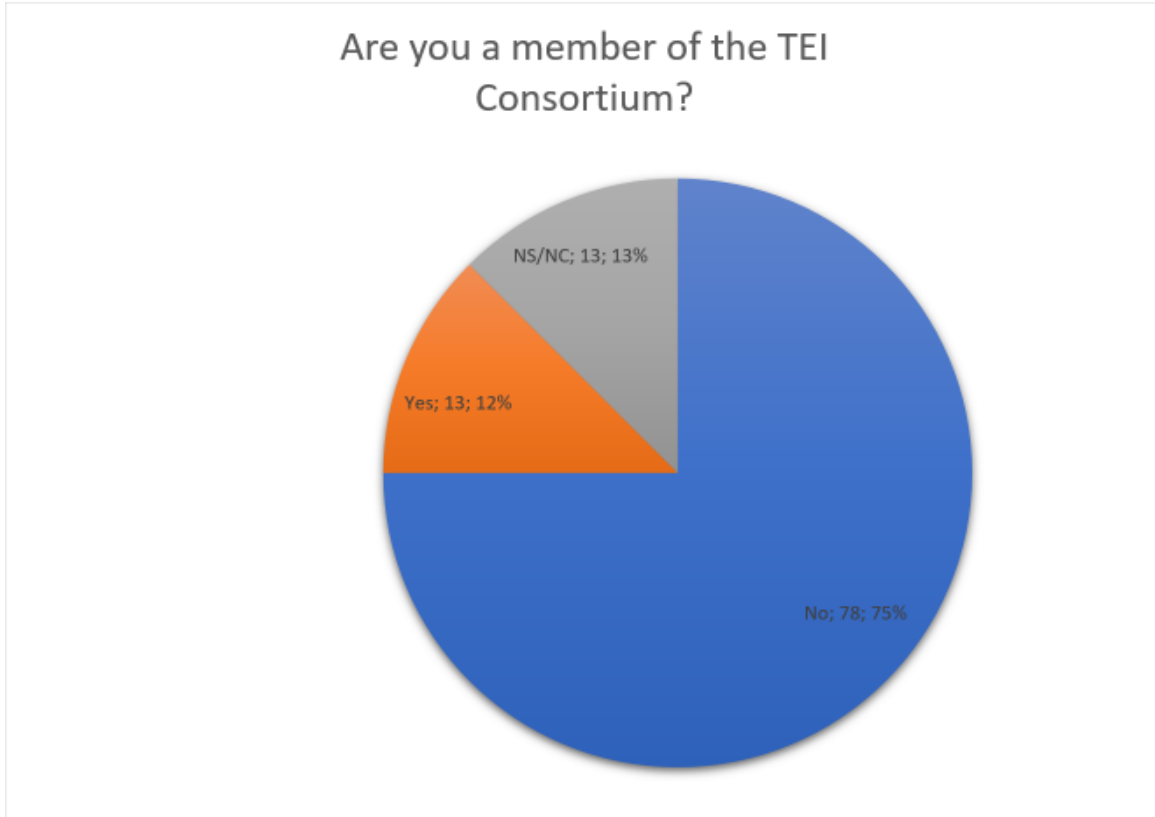


Figure 12. Consortium Membership: Institutions

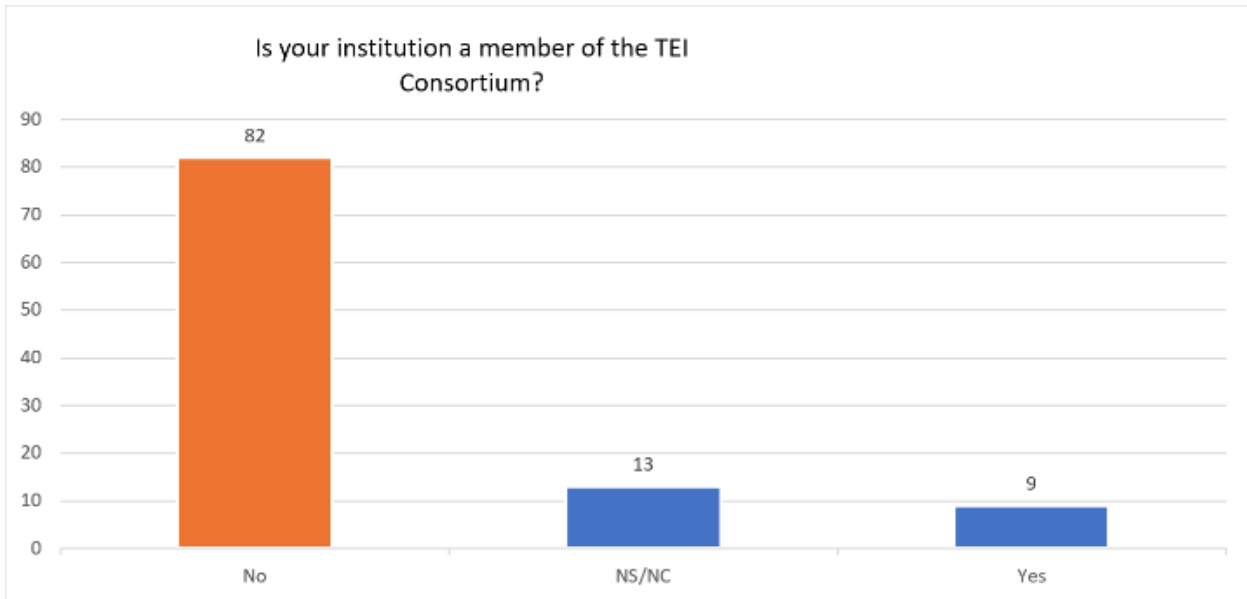


Figure 13. Participation in TEI Conferences

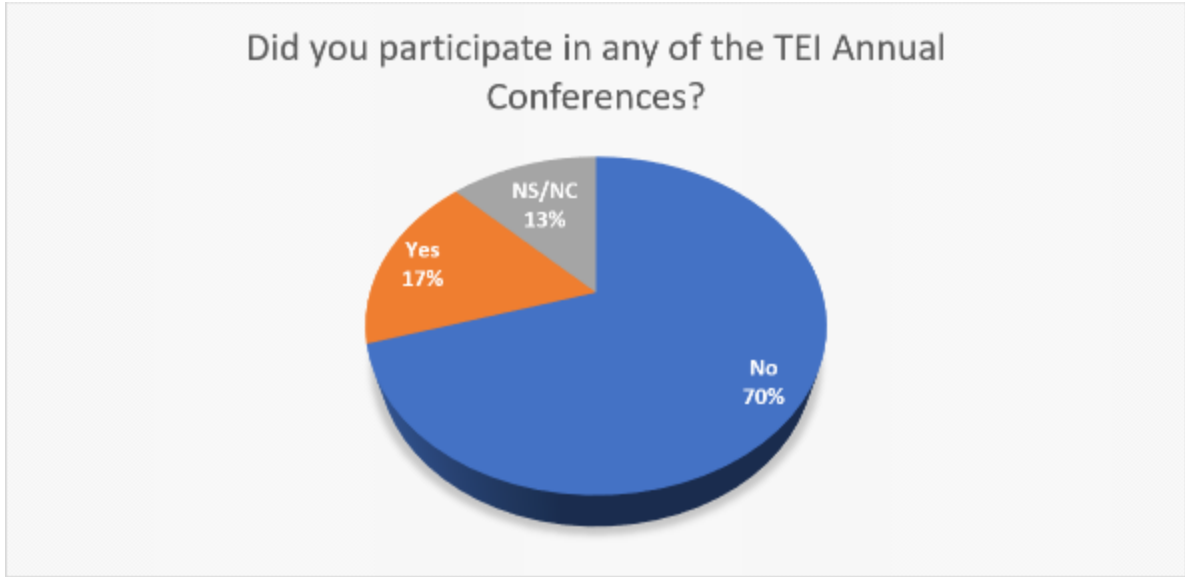


Figure 14. Platforms and Tools used to transform and publish TEI files

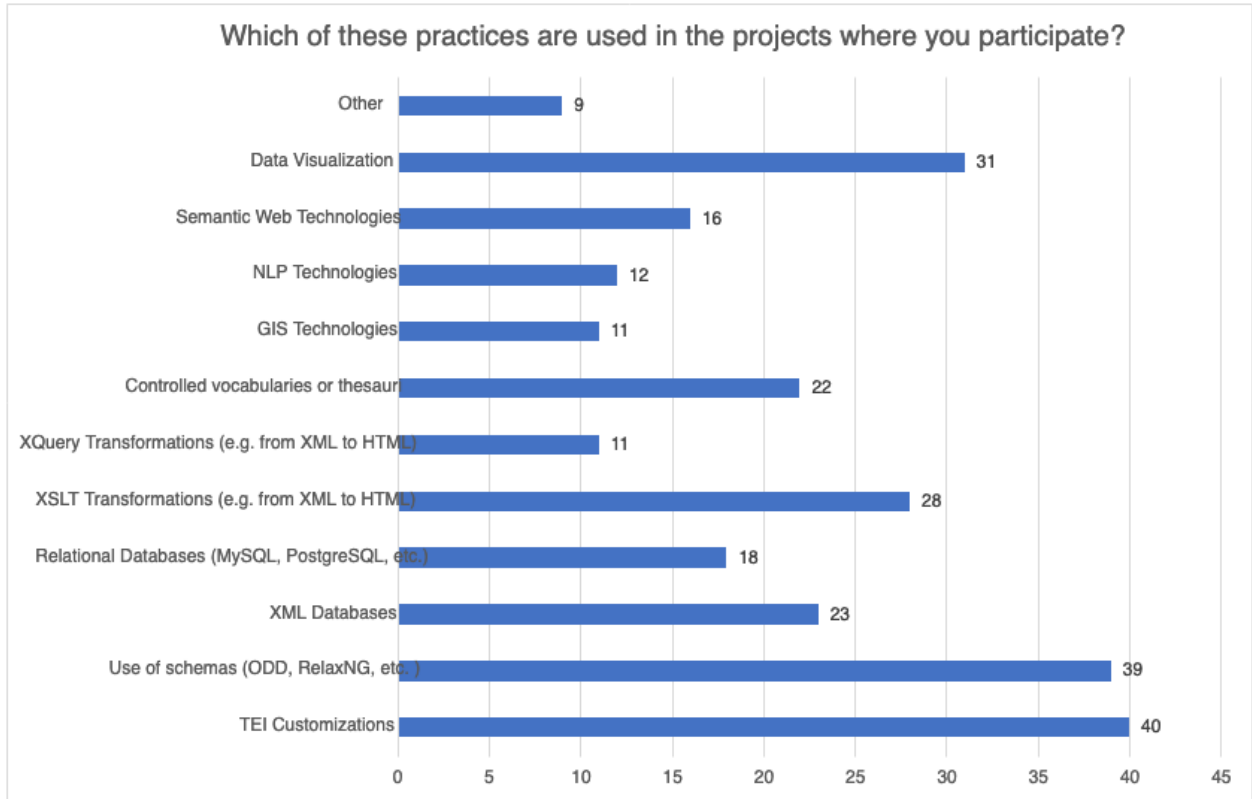


Figure 15. TEI Encoded File Sharing

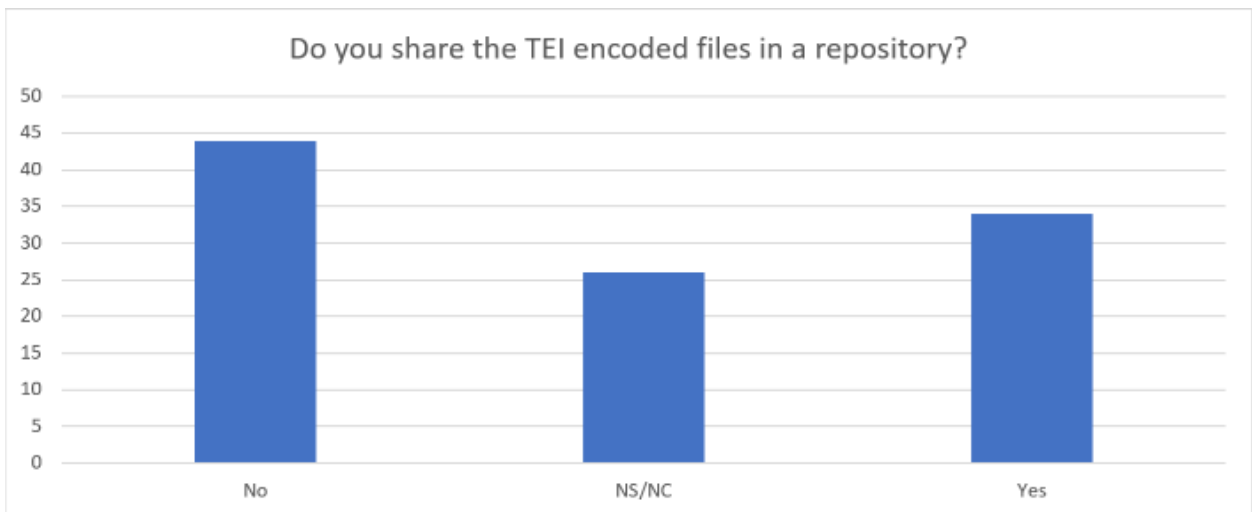


Figure 16. Platforms or tools used to transform or publish TEI files

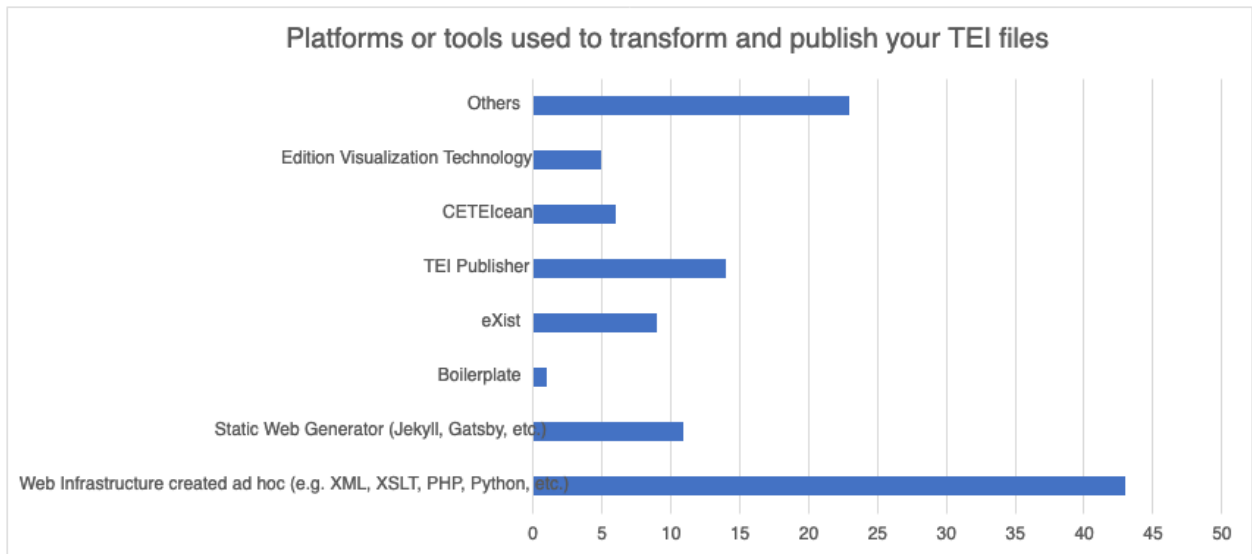


Figure 17. Levels of TEI Teaching

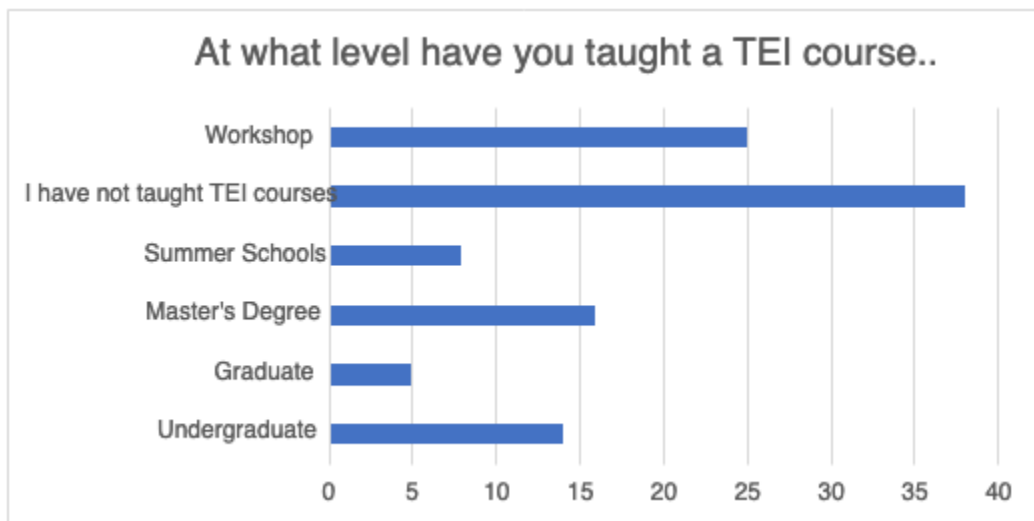


Figure 18. Resources for Teaching TEI

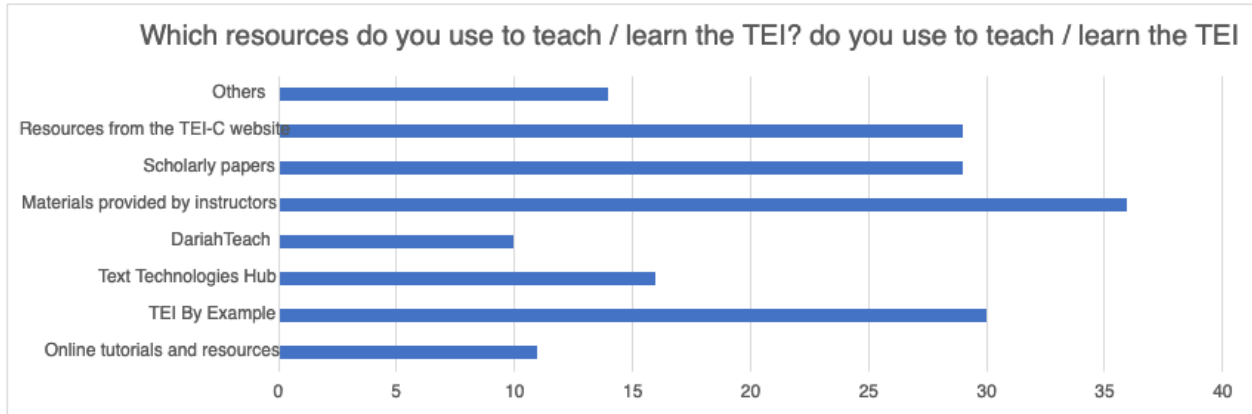
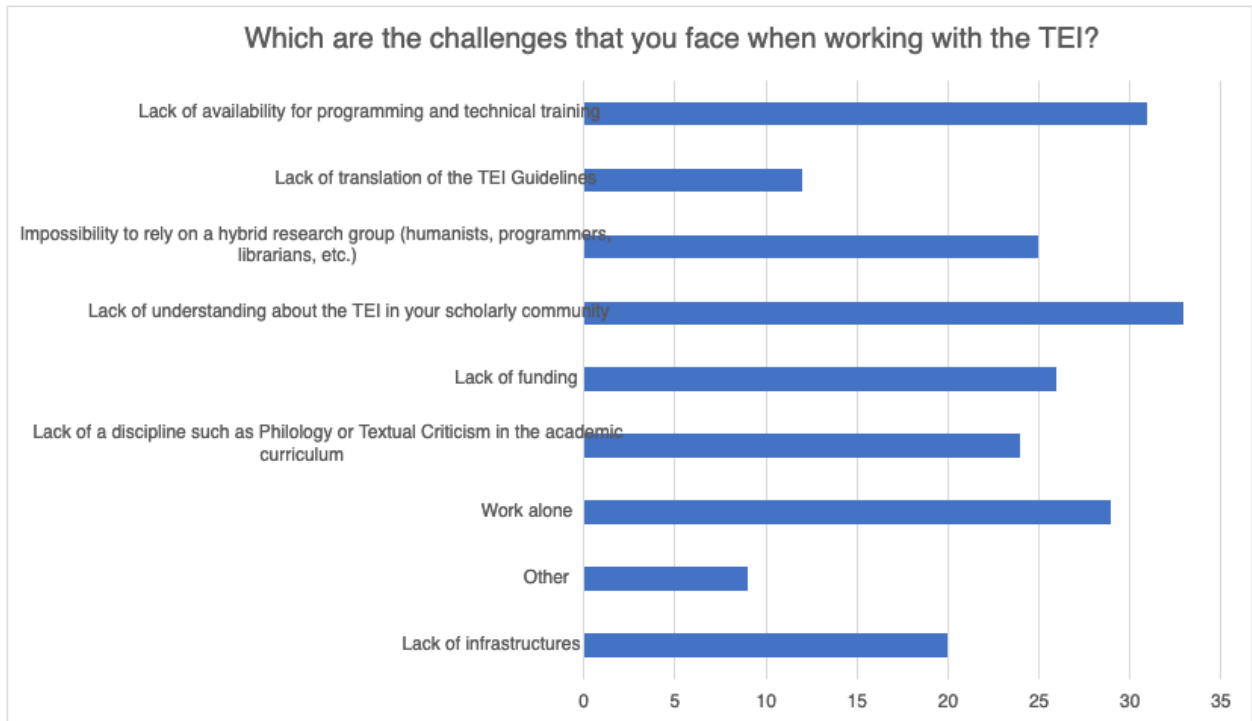


Figure 19. Challenges when working with the TEI



Appendix 5. List of Projects for Q11

Number	Name	URL Provided
1	Archive of Biographical Writings in Medieval and Early Modern Iberian	https://archbio.miami.edu/
2	Schopenhauer's Library. Annotations and marks in his Spanish books	http://www.schopenhauer.uni.wroc.pl
3	Biblioteca Digital EMOTHE	https://emothe.uv.es/
4	ColoniaLab. A laboratory for the collaborative digital editing of materials related to colonial Latin America	https://colonialab.org/
5	El «Libro de Apolonio». Una edición crítica digital	https://doi.org/10.5167/uzh-206040
6	Proyecto de desarrollo de tecnología educativa	http://educafaum.weebly.com/
7	Fairsfair. Fostering Fair Data Practices in Europe	https://www.fairsfair.eu/
8	P.S. Post Scriptum. A Digital Archive of Ordinary Writing (Early Modern Portugal and Spain)	http://teitok.clul.ul.pt/postscriptum/
9	Estoria de Espanna Digital	Estoria.bham.ac.uk
10	ANR D4R : Dissidences religieuses et réception de la Réforme à la Renaissance (Espagne-XVIe s.)	https://iriec.www.univ-montp3.fr/fr/valorisation-partenariats/presse/projets-de-recherche/anr-d4r-dissidences-religieuses
11	Der arme Heinrich – digital	https://digi.ub.uni-heidelberg.de/en/ahd/index.html
12	Edição Digital de Fernando Pessoa. Projetos e Publicações	http://www.pessoadigital.pt/pt/index.html
13	Digitale Briefedition Alfred Escher	https://escher.sources-online.org/home.html
14	DraCor	https://dracor.org/

15	Diccionario Griego-Español (DGE)	http://dge.cchs.csic.es/
16	Oscar Wilde Digital	https://semillero poeticaytraduccion.com/
17	Bieses. Biografías de Escritoras Españolas	https://www.bieses.net/
18	Léxico Leonés Actual	https://lla.unileon.es/
19	Corpus del Español del Siglo XXI (CORPES)	https://apps2.rae.es/CORPES/
20	Cartas a la familia. De la migración de Jesusita a Jane	https://familyletters.unl.edu/
21	Poesía Medieval	http://hdlab.space/Poesia-Medieval/
22	Biblioteca Electrónica Textual del Teatro en Español (BETTE)	https://github.com/HDAUNIR/BETTE
23	Biblioteca virtual. Prueba de Concepto sobre el legado Luciano Pereña	http://www.larramendi.es/luc_per/es/micrositios/inicio.do
24	DraCor - Spanish Drama Corpus	https://www.dracor.org/span
25	Avisos de Levante	https://avisosdelevante.wordpress.com
26	Amadis in Translation	https://amadis.newtfire.org/amadis/
27	Clásicos Hispánico	www.clasicoshispanicos.com/
28	El mercado digital de texto : una exploración con el archivo periodístico sobre desapariciones forzadas ocurridas en el conflicto armado interno del Perú	http://hdl.handle.net/1992/48466
30	The SCTA Reading Room. A site for reading, viewing, and studying the scholastic tradition	https://scta.lombardpress.org/
31	Nicolasa de Helguero y Alvarado. Edición digital de la obra poética	https://proyectobieses.github.io/helguero/
32	Biblioteca del Pensamiento NovoHispano	www.bdpn.unam.mx
33	Thebarum Fabula. Biblioteca digital del mito tebano con ediciones críticas y traducciones	http://thebarumfabula.usc.es/en/

34	CELT, the Corpus of Electronic Texts	http://research.ucc.ie/celt/
35	Proyecto Mambrino	https://www.mambrino.it/it
36	Bidiso. Biblioteca Digital del Siglo de Oro	https://www.bidiso.es/index.htm
37	Comedic. Catálogo de obras medievales impresas en castellano	https://comedic.unizar.es/
38	La Entretenida by Miguel de Cervantes	http://entretenida.outofthewings.org/
39	The School of Salamanca. A Digital Collection of Sources and a Dictionary of its Juridical-Political Language	https://www.salamanca.school/en/index.html
40	Distant Reading for European Literary History	https://www.distant-reading.net/
41	7 Partidas Digital. Edición crítica de las Siete Partidas	https://7partidas.hypotheses.org/
42	tei-ausiasmarch	https://doi.org/10.5281/zenodo.5091724
43	The Chymistry of Isaac Newton	https://webapp1.dlib.indiana.edu/newton/
44	Latin American Comics Archive	http://mlrdev.hss.cmu.edu/omeka/

Appendix 6. List of projects providing their XML-TEI files

Number	Name	URL Provided
1	Hub de Tecnologías del Texto	https://github.com/tthub-repo/ejemplos
2	Estoria Digital	https://zenodo.org/record/2593589#.YgKT1erP3IV
3	Hartmann's von Aue <i>Der arme Heinrich</i>	https://heidata.uni-heidelberg.de/dataset.xhtml?persistentId=doi:10.11588/data/FNKJ2Y
4	Digital Edition of Fernando Pessoa	https://github.com/cceh/pessoa
5		https://github.com/eeditiones/dodis-wall
6	Dracor	https://dracor.org/

7		https://www.eafit.edu.co/biblioteca/sala-patri-monio-documental/Paginas/ediciones-digitales.aspx
8		https://github.com/orgs/CDRH/teams/family-letters/repositories
9		http://hdlab.space/Poesia-Medieval/
10	Humbolt Digital	https://github.com/humboltdigital/ediciones
11		https://repositorio.linguisticafeminista.com/home
12	La Biblioteca Electrónica Textual del Teatro en Español (BETTE)	https://github.com/HDAUNIR/BETTE
13		https://github.com/HDAUNIR/BETTE
14	Poesía Medieval	https://github.com/hdcaicyt/Poesia-Medieval
15	Amadis in Translation	https://github.com/ebeshero/Amadis-in-Translation
16	Clásicos Hispánicos	https://zenodo.org/communities/clasicos_hispanicos?page=1&size=20
17	Family Letters	https://github.com/CDRH/data_family_letters/tree/dev/source/tei
18	Notas periodísticas	https://github.com/lgarciabe/notas_periodisticas
19		https://github.com/scta-texts/
20	Bieses - Helguero	https://github.com/ProyectoBIESES/helguero
21		http://thebarumfabula.usc.es/exist/apps/eXide/index.html
22		https://github.com/DesenrollandoElCordel
23		http://research.ucc.ie/celt/
24		https://github.com/digicademy/svsal-tei
25	ELTeC	https://github.com/COST-ELTeC
26	7 Partidas Digital	https://github.com/7PartidasDigital/XML-TEI
27	EMOTHE	https://emothe.uv.es/

28		https://github.com/Cantavestrella/tei-ausiasma_rch
29	Biblioteca del Pensamiento NovoHispano	http://www.bdpn.unam.mx/

Appendix 7. List of Activities

1. Roundtable at *Scire Vias Humanidades Digitales y Conocimiento*. V Congreso de la HDH. Santiago de Compostela, Spain. 4 - 8 de octubre de 2021. <https://hdh2021.org/>

Panel

Experiencias en torno a la Text Encoding Initiative: Retos de codificación, plataformas de publicación y creación de corpus

Palabras clave: Text Encoding Initiative, edición digital, etiquetado de textos, anotación semántica

Coordinadora: Susanna Allés Torrent, University of Miami

Participantes:

Antonio Rojas Castro, Berlin-Brandenburgische Akademie der Wissenschaften y Universidad Internacional de la Rioja
Leyre Martín Aizpuru y Víctor Caballero Gómez, Universidad de Sevilla
Marta López Izquierdo, Université Paris 8

Este panel presenta diversas iniciativas relacionadas con el uso de la Text Encoding Initiative (TEI) con el objetivo de reflexionar sobre los retos afrontados a lo largo de su adopción y aplicación: ya sea en su concepción, el modelado de los datos, la creación de infraestructuras para la publicación, la colaboración en el etiquetado de los textos, o incluso en la exploración de los datos textuales desde un punto de vista cuantitativo y translingüístico. La sesión promueve un diálogo constructivo que ilumine los escollos y las carencias del uso de TEI en el seno de la comunidad de usuarios en español, al mismo tiempo que evidencie los nuevos derroteros por los que esta práctica debe discurrir.

La TEI es una de las iniciativas académicas en el ámbito de las Humanidades Digitales más conocidas. Su trayectoria, ahora ya con más de treinta años, y una comunidad global con miles de usuarios, hace de esta iniciativa una de las más sólidas para la codificación informática de textos humanísticos y del ámbito de las Ciencias Sociales. Aunque su uso y su sistema de marcado se ha aplicado especialmente en el campo de la edición de textos, independientemente de la escuela o tradición de crítica textual, cada vez son más los escenarios en los que se adopta TEI, como puede ser la creación de corpus lingüísticos e incluso para la estructuración de contenidos textuales a modo de base de datos.

La TEI se ampara en un consorcio compuesto por personas e instituciones, principalmente norteamericanas y europeas. Este consorcio es el responsable de desarrollar y mantener actualizadas las Guías directrices (*Guidelines*) para el marcado (también llamado etiquetado y codificación) de los textos en formato digital. Estas guías son una especie de manual de codificación o de recomendaciones de buenas prácticas que especifican una metodología concreta de marcado de los textos de manera que puedan ser procesados informáticamente⁷². El

⁷² <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html>

espíritu de acceso abierto y colaboración del consorcio hace que sean los mismos usuarios los que tengan la capacidad para proponer mejoras y modificaciones a través de su repositorio en GitHub. Este espíritu de colaboración y de difusión se manifiesta también a través de un diálogo constante a través de la lista de discusión⁷³ y de la organización anual de una conferencia. El consorcio, en fin, crea grupos de trabajo y de interés para ciertas disciplinas, como pueden ser diccionarios, lingüística de corpus, descripción de manuscritos, etc. Todo ello hace que la comunidad TEI sea cada vez más amplia, y al mismo tiempo que más especializada.

En el ámbito del español, la TEI se ha utilizado y estudiado desde finales de los ochenta. Entre las publicaciones pioneras se encuentra las de Faulhaber (1994), Marcos Marín (1986, 1994) o Irizarry (1997), que entendieron ya con precisión las potencialidades de un estándar que podía crear textos enriquecidos, interoperables, preservables y con funcionalidades de búsqueda semántica. A lo largo de los 2000 también fueron apareciendo, aunque tímidamente, escritos divulgativos o centrados en proyectos concretos de investigación (Fradejas Rueda, 2009; o los múltiples trabajos de Alex Bias). También aparecieron las primeras publicaciones sobre las posibilidades de la edición digital académica, que desde entonces han tenido una cierta continuidad (Lucía Megías, 2008). Y más recientemente, han visto la luz otros artículos de alcance más general y centrados sobretudo en España (Spence, 2014; Allés Torrent, 2017; Rojas Castro, 2017), o en casos de estudio de proyectos en curso (Revenga, 2014; González-Blanco et al., 2014; González- Blanco & Rodríguez, 2015; del Río Riande & Zubillaga, 2015; Navarro Colorado 2015; Rojas Castro, 2017). En lo que respecta específicamente a TEI, y sus especificaciones técnicas la literatura en español es todavía poco extensa, aunque esta tendencia va mejorando significativamente (Fradejas, 2009; Allés Torrent, 2015; Rojas Castro, 2018; Schöch et al. 2019).

De manera paralela, fueron surgiendo también desde los años 90 diferentes proyectos que adoptaron TEI como método de marcado de textos. Algunos de los ejemplos más conocidos son la Biblioteca Virtual Miguel de Cervantes (1999), que ofrecía facsímiles con imágenes digitales con cierta calidad. También los corpus del CREA y del CORDE de la Real Academia de la Lengua Española llevó a cabo una codificación y ejecución similar. En ambos casos, desafortunadamente, nunca pusieron a disposición de los usuarios los archivos XML-TEI en abierto para que toda la comunidad pudiera usar como referencia o tratar con ellos como lo que eran, datos informáticos para ser procesables. Desde hace unos años, la comunidad internacional ha puesto de relieve la necesidad de compartir los archivos codificados y esta práctica empieza a ser también adoptada por los proyectos en español. A día de hoy, empiezan a ser más numerosos los ejemplos tanto de ediciones digitales como de corpus lingüísticos que, además de una documentación sobre el tipo de marcado, ofrecen en abierto los archivos fuente en XML-TEI⁷⁴. Entre los primeros, y especialmente fértil entre los estudios del Siglo de Oro (Allés Torrent, 2017) encontramos los proyectos editoriales de *La Dama Boba* (ed. Prolope, 2015)⁷⁵, las *Soledades* de Luís de Góngora (ed. Rojas Castro, 2016)⁷⁶, el corpus de Poesía medieval (2020)⁷⁷ o las *Siete Partidas Digital* (ed. Fradejas Rueda, 2018)⁷⁸. Entre los corpus, también se encuentran

⁷³ Más información sobre la lista de discusión está disponible en: <https://tei-c.org/support/>

⁷⁴ Véase una lista con ejemplos sobretudo en español, en <https://tthub.io/recursos/ejemplos-tei/>

⁷⁵ <http://damaboba.unibo.it/>

⁷⁶ <http://soledades.uni-koeln.de/>

⁷⁷ <http://hdlab.space/Poesia-Medieval/>

⁷⁸ <https://7partidas.hypotheses.org/>

iniciativas sumamente relevantes como el CLiGS textbox⁷⁹ o DRACOR⁸⁰, de carácter internacional y multilingüe y que contienen una sección de corpus etiquetados en español.

La relativa lentitud en la adopción de TEI para proyectos en español se debe también a la escasez de materiales para su aprendizaje. A diferencia de la comunidad anglófona, que desde hace ya más de diez años tiene tutoriales en línea, como el TEI By Example⁸¹, no ha ocurrido así en nuestro caso. Recientemente se han traducido tutoriales de TEI del inglés (Schreibman et al., 2020), pero el reto es todavía el de la creación de materiales que realmente se adapten a la tradición cultural y tomen ejemplos concretos en esta lengua. Con tal fin, han aparecido iniciativas como el *Text Technologies Hub. Recursos sobre tecnologías del texto y edición digital* (TTHub)⁸² que ofrece lecciones en línea, de diferente nivel, una bibliografía sobre TEI en español, e incluso una base de datos *in fieri* de los proyectos que ofrecen ejemplos concretos de marcado.

Apostando por la diversidad lingüística y cultural, el Consorcio (TEI-C) está explorando actualmente estrategias para aumentar la accesibilidad para aquellos usuarios que trabajan con lenguas diferentes del inglés. Para ello ha creado un grupo de trabajo para la internacionalización (*Internationalization Working Group*) que, entre otros objetivos, tiene el de la creación de un glosario multilingüe de términos utilizados en las guías directrices de TEI y un repositorio de ejemplos en lenguas y orígenes diferentes. Este trabajo irá acompañado de la reestructuración de la página web para que pueda dar una mayor visibilidad a los contenidos y a la dimensión internacional del Consorcio, como por ejemplo, páginas de bienvenida en diferentes lenguas. Otra de las líneas de trabajo es la redacción de materiales introductorios en diferentes lenguas para acercar más fácilmente a los usuarios de otras comunidades lingüísticas, así como el desarrollo de una infraestructura para hacer la traducción de las especificaciones de TEI más ágil. Como es sabido, tanto el sitio de la TEI, como sus guías directrices solo son consultables en inglés, y a día de hoy la traducción en español sigue siendo solo parcial⁸³.

Bajo el auspicio de esta iniciativa del Consorcio TEI, financiada este año por la fundación Mellon, este panel pretende analizar el uso y la vitalidad del mercado con TEI en español. En esta ocasión, se agrupan experiencias de investigadores españoles, activos en diferentes centros nacionales e internacionales. Cada una de las intervenciones trata un tema clave para el desarrollo del uso de TEI en español con diferentes niveles de familiaridad y de complejidad. La primera intervención (Rojas Castro) aborda el tema crucial de la infraestructura y la publicación de los archivos en TEI y trae a colación el problema lingüístico en la interfaces web. Las otras dos presentaciones se centran en la creación de corpus textuales a partir del estándar propuesto por la TEI y con materiales textuales de los siglos XIX y XX. Ambos proyectos, CODHECUN y CAREXIL-FR, ofrecen un corpus bien delimitado y han optado por el uso de la plataforma TEITOK, una plataforma web para la visualización, creación y edición de corpus con marcado semántico (como XML-TEI) y anotación lingüística. Estas cuatro iniciativas pioneras ofrecerán una visión de conjunto del estado actual sobre edición digital y creación de

⁷⁹ <https://github.com/cligs/textbox>

⁸⁰ <https://dracor.org/span>

⁸¹ <https://teibyexample.org/>

⁸² <https://tthub.io/>

⁸³ Verónica Zumárraga y Marcela Tabanera tradujeron una primera versión de las guías directrices de la TEI en su *Manual para la codificación e intercambio de textos informatizados. Normas de la Text Encoding Initiative* (Marcos Marín, 1994: 89). Más tarde, Alejandro Bia y Carmen Arronis Llopis empezaron la traducción de las guías directrices en su versión P5 pero estas permanecen inconclusas.

corpus y abren las fronteras para que los humanistas digitales puedan finalmente transitar y aprovechar al máximo los datos producidos a través de herramientas de análisis textual.

Ich verstehe gar nichts... Sobre la localización de ediarum al español

Antonio Rojas Castro

Ediarum es un framework de edición en formato TEI diseñado e implementado por TELOTA en la Academia de Ciencias y Humanidades de Berlín-Brandeburgo (BBAW, Alemania) basado en dos componentes principales: una base de datos XML de código abierto (eXistDB) y un editor XML comercial ampliamente utilizado (Oxygen XML editor). Estos dos componentes se comunican mediante una interfaz en alemán utilizada en proyectos de Humanidades Digitales desarrollados en la Academia de Berlín. En 2020 el proyecto de cooperación cubano-alemán Humboldt Digital (ProHD) emprendió un proceso de adaptación de la interfaz de ediarum con el objetivo de facilitar la tarea de edición de textos en formato TEI: por un lado, se desarrollaron las funciones y el código para que fueran independientes de la lengua (internacionalización) y, por el otro, se localizó la interfaz al español mediante la creación de un fichero con las traducciones (localización). Como resultado de este proceso, Proyecto Humboldt Digital ha desarrollado un nuevo framework llamado ediarum.PROHD.edit disponible en Github. El objetivo de esta contribución es presentar ediarum.PROHD.edit y reflexionar sobre los retos más importantes encontrados durante la localización.

Corpus Documental y Hemerográfico de la Cuba del Novecientos (CODHECUN): diseño, elaboración y explotación de un corpus en la plataforma TEITOK

Leyre Martín Aizpuru y Victor Caballero Gómez

En el marco del proyecto “Cuba y Andalucía en el siglo XIX: estudio de los lazos lingüísticos y culturales desde las Humanidades Digitales” (<https://institucional.us.es/cuba19/>), estamos realizando la transcripción y edición digital de un corpus de acuerdo con los criterios de la Red CHARTA en su versión digital (Isasi *et alii*, 2020) y en el marco de la plataforma TEITOK, desarrollada por Maarten Janssen (2014). Además de ofrecer este corpus tokenizado, normalizado, lematizado y anotado morfosintácticamente, nos hemos propuesto marcar aspectos léxico-semánticos, con el propósito de elaborar índices temáticos. Ello nos permitirá poner a disposición de otros investigadores los documentos del corpus, así como llevar a cabo un análisis de los usos lingüísticos y discursivos del español del siglo XIX, tomando Cuba como referencia geográfica y centro del espacio variacional caribeño, con el propósito de definir la especificidad del español cubano y sus lazos con la sociedad y la lengua andaluza del Novecientos.

TEI para un proyecto transdisciplinar: construyendo CAREXIL-FR

Marta López Izquierdo

El proyecto en construcción [CAREXIL-FR](https://carexil.univ-paris8.fr/) (CARTas de REpublicanos Españoles REfugiados y EXILIados en FRancia, <https://carexil.univ-paris8.fr/>) nace como un empeño transdisciplinar de edición y estudio de un corpus epistolar inédito, para responder a las necesidades de un equipo que reúne a filólogos, lingüistas, historiadores, cartógrafos, antropólogos, especialistas en

humanidades digitales, en literatura y en arte. Por un lado, hemos desarrollado un protocolo de transcripción, edición y anotación basado en los estándares TEI, siguiendo la pauta de otros proyectos de corpus epistolares digitales (Post-scriptum), pero adaptándola a las necesidades del estudio transdisciplinar específico de CAREXIL-FR. Por otro lado, la adopción del entorno TEITOK (Maarten Janssen, 2014) conlleva sus propias normas de modelado y anotación de los documentos que es necesario integrar en la fabricación del corpus. En nuestra intervención, expondremos algunos de los escollos que hemos encontrado a lo largo de nuestro trabajo y las soluciones que estamos proyectando.

Bibliografía

- Allés Torrent, S., y del Rio Riande, G. (2018). Enseñar edición digital con TEI en español. Aprendizaje situado y transculturación. En *Humanidades Digitales: construcciones locales en contextos globales*. Actas del I Congreso Internacional de la Asociación Argentina de Humanidades Digitales (AAHD), Buenos Aires: Editorial de la Facultad de Filosofía y Letras Universidad de Buenos Aires.
- Allés Torrent, S. (2015). Edición digital y algunas tecnologías aliadas. *Insula*, 822: 18–21.
- Allés Torrent, S. (2017). Tiempos hay de acometer y tiempos de retirar: Literatura áurea y edición digital. *Studia Aurea*, 11: 13–30.
- Faulhaber, Ch. (1994). La Text Encoding Initiative y su aplicación a la codificación textual y explotación. En *Actas del Congreso de la lengua española: Sevilla, 7 al 10 Octubre, 1992*, 331–40. Madrid: Instituto Cervantes.
https://datospdf.com/download/la-text-encoding-initiative-y-su-aplicacion-a-la-codificacion-textual-y-explotacion-_5a44beefb7d7bc422b88bbd1_pdf.
- Fradejas Rueda, J. M. (2009-10). La Codificación XML/TEI de Textos Medievales. *Memorabilia*, 12: 219–47.
<https://parnaseo.uv.es/Memorabilia/Memorabilia12/PDFs/Codificacion.pdf>
- González-Blanco, E., y Rodríguez, J. L. (2014). ReMetCa: A Proposal for Integrating RDBMS and TEI-Verse. *Journal of the Text Encoding Initiative*, 8.
<https://doi.org/10.4000/jtei.1274>.
- Irizarry, E. (1997). *Informática y Literatura. Análisis de Textos Hispánicos*. Madrid: Anthropos.
- Lucía Megías, J. M. (2008). Las relaciones entre la bibliografía textual y la informática humanística: el incunable del hipertexto. *Tipofilología. Rivista Internazionale di Studi Filologici e Linguistici sui testi a stampa*, 1: 119–38.
- Marcos Marín, F. (1986). Metodología informática para la edición de textos. *Incipit*, 6: 185–97.
- Marcos Marín, F. (1994). *Informática y Humanidades*. Madrid: Gredos.
- Navarro Colorado, B. (2015). A Computational Linguistic Approach to Spanish Golden Age Sonnets: Metrical and Semantic Aspects. En *Proceedings of the Fourth Workshop on Computational Linguistics for Literature*. Denver.
http://www.dlsi.ua.es/~borja/navarro2015_GoldenAgeSonnets.pdf.
- Revenge García, N. (2014). La edición crítica digital de textos teatrales. La estrella de Sevilla como ejemplo. *Anuario de Lope de Vega. Texto, Literatura, Cultura*, 20: 99–121.
- Rio Riande, G., y Zubillaga, C. (2015). Un proyecto de edición digital académica en Argentina. Diálogo Medieval. En *Actas I Jornadas de Humanidades Digitales de la AAHD*. Buenos

- Aires: Editorial de la Facultad de Filosofía y Letras.
<https://www.aacademica.org/gimena.delrio.riande/15>.
- Rojas Castro, A. (2018). La edición académica digital. De las teorías del texto a la visualización de la información. En I. Galina Russell, J. F. Barrón Tovar, M. Peña Pimentel, E. Priani Saisó, D. Domínguez Herbón, and A. Álvarez Sánchez (Eds.), *Humanidades Digitales: Edición, Literatura y Arte*. CDMX: Bonilla Artigas.
- Rojas Castro, A. (2017). La edición crítica digital y la codificación TEI. Preliminares para una nueva edición de las *Soledades* de Luis de Góngora. *Revista de Humanidades Digitales*, 1.0: 4–19. <https://doi.org/10.5944/rhd.vol.1.2017.16379>.
- Schöch, C., Calvo Tello, J., Henny-Krahmer, U., y Popp, S. (2019). The CLiGS Textbox: Building and Using Collections of Literary Texts in Romance Languages Encoded in TEI XML. *Journal of the Text Encoding Initiative*. <https://doi.org/10.4000/jtei.2085>.
- Schreibman, S., y Bleier, R. (2020). La codificación textual y la Iniciativa TEI. *dariahTeach. DH Teaching material*. <https://teach.dariah.eu/mod/page/view.php?id=739>
- Spence, P. (2014). Edición académica en la era digital: modelos, difusión y proceso de investigación. *Anuario Lope de Vega. Texto, literatura, cultura*, 20.0: 47–83. <https://doi.org/10.5565/rev/anuariolopedevega.74>.

2. Events at Next-Gen TEI TEI Conference and Members' Meeting (Online). October 25-30, 2021

2.1. Roundtable: TEI en Español

Chair: Gimena del Rio Riande, Consejo Nacional de Investigaciones Científicas y Técnicas

- Gimena del Rio Riande, Susanna Allés Torrent - [TEI en español: Why and How](#)
- Ernesto Priani - Teaching TEI from distance and in Spanish. Advantages and limits
- Gabriel Alejandro Calarco^{1,2} - The ekphrasis in the *Libro de Alexandre*, a digital publishing project on the study of the Castilian clerical poetry in the thirteenth century
- Cristian Suárez Giraldo - Edición académica e investigación formativa en literatura. Usos de XML-TEI en el pregrado de Literatura de la Universidad EAFIT

Abstract:

2.2. Short Paper: [i18n: Building a TEI Multilingual Community](#)

Participants: Susanna Allés-Torrent, Hugh Cayless, Gimena del Rio Riande, Luis Meneses, Kiyonori Nagasaki, Martina Scholger, and many more collaborators!

I18n: Building a TEI Multilingual Community

Keywords: translation, multilingualism, community-building, Text Encoding Initiative, internationalization

The Text Encoding Initiative [1] is one of the oldest and consolidated DH communities and one of the most continuously developed DH projects in existence. Although efforts have been made to translate the TEI Guidelines and the specifications into languages other than the original English [2][3][4], it is difficult to stay continuously up to date with the Guidelines, which change incrementally and employ terminology that is often technical and sometimes unclear. The TEI working group on internationalization (i18n) was established in 2019 with the main goals of focusing on the challenge of communicating the TEI in different languages and developing proposals to make the TEI a more multilingual, international and open community [5][7].

Language is a primary medium in the transmission of culture and ideas. The flow of knowledge as we try to communicate the TEI cannot be simply directed from English to other languages. Specifically, direct translations can introduce concepts to the Guidelines that are incorrect or unclear.

i18n is currently developing a glossary of TEI technical terms, accompanied by notes and examples for translators. Additionally, the TEI's website is being redesigned to facilitate internationalization efforts and address a multilingual audience. Most recently, and with the support of the Andrew Mellon Foundation, team members are working towards the development of a translation tool and workflow in order to smoothe and overcome the hurdles when translating a massive and specialized corpus such as the TEI specifications. Also, the team is engaged with the improvement and creation of original teaching materials in other languages (German, Japanese), such as the Text Technologies Hub [2], which currently offers a basic introduction to the TEI.

Although it is probably inevitable that English be used as a common basis of discourse for the TEI, a concerted effort must be made to provide introductory materials in other languages [6][2]. This presentation will document the steps we are taking towards ensuring that communication between the different linguistic communities using TEI can flow in multiple directions. The final goal is to create more open and equitable channels through which the TEI is better communicated to a more diverse user base.

References

- [1] TEI Consortium, eds. *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. 4.2.0. February 2021. TEI Consortium. <http://www.tei-c.org/Guidelines/P5/> (2021-08-30).

- [2] Allés Torrent, Susanna / Calarco, Gabriel / Rio Riande, Gimena del. TTHub. Text Technologies Hub. Recursos sobre tecnologías del texto y edición digital: <https://tthub.io/> (2021-08-30).
- [3] Stadler, Peter / Scholger, Martina (2016): Report on the TEI2German Translatathon at Vienna 2016. <https://tei-c.org/tools/i18n/tei2german> (2021-08-30).
- [4] SIG East Asian/Japanese, TEI Consortium: <https://github.com/TEI-EAJ> (2021-08-30).
- [5] Bermúdez Sabel, Helena / Cayless, Hugh / Meneses, Luis / Nagasaki, Kiyonori / Rio Riande, Gimena del / Scholger, Martina (2019):
<https://triangleosci.org/2019/07/17/communicating-the-tei-to-a-multilingual-user-community>
(2021-08-30).
- [6] Terras, Melissa / Van den Branden, Ron / Vanhoutte, Edward (2017): TEI by Example.
<http://www.teibyexample.org> (2021-08-30).
- [7] Cayless, Hugh / Bermúdez Sabel, Helena / Meneses, Luis / Rio Riande, Gimena del / Nagasaki, Kiyonori (2020): Communicating the TEI Across Linguistic and Cultural Boundaries. DH2020 Conference.
https://dh2020.adho.org/wp-content/uploads/2020/07/528_CommunicatingtheTEIAcrossLinguisticandCulturalBoundaries.html (2021-08-30).

3. Workshop “La Text Encoding Initiative y el lenguaje XML: introducción al marcado semántico de textos”, at the CCHS-CSIC, Madrid (Spain), January 31 - February 1, 2022.

SEMINARIOS TRANSVERSALES DEL CCHS

**LA TEXT ENCODING INITIATIVE Y EL LENGUAJE XML:
INTRODUCCIÓN AL MERCADO SEMÁNTICO DE TEXTOS**

DOCENTE: SUSANNA ALLÉS-TORRENT
UNIVERSITY OF MIAMI

Fecha de celebración: 31 de enero y 1 de febrero de 2022
Horario: 10–13h
Taller virtual

Descripción: Este seminario se concibe como una introducción genérica al lenguaje de marcado Extensible Markup Language (XML) y a las guías directrices de la Text Encoding Initiative (TEI).

XML-TEI es un marco de trabajo concebido especialmente para la codificación, la explotación informática, la publicación en línea y la preservación de textos en ciencias sociales y humanidades, utilizado internacionalmente y por una gran comunidad de usuarios.

XML-TEI ofrece una gran variedad de campos de aplicación, no solo en proyectos relacionados con la literatura o la lingüística, sino también en instituciones culturales tales como bibliotecas, museos, archivos e incluso editoriales.

El seminario se dividirá en dos sesiones de 3 horas y responde a las necesidades de los investigadores, técnicos y responsables de proyectos del CCHS. Cada una de las sesiones se estructurará en una presentación teórica, seguida de sesiones prácticas con ejercicios concretos para dar una visión de conjunto de la metodología de trabajo necesaria para adoptar este tipo de anotación semántica.

4. “TEI Editaton,” SemanaHD, organized by Asociación Argentina de Humanidades Digitales, la Red Colombiana de Humanidades Digitales and la Red HD de México, 13 al 19 June 13th-19th, 2022.

Abstract

The presentation is uploaded in our Zenodo community: Allés Torrent, Susanna, Gimena del Rio Riande, & Hugh Cayless. (2022, June 17). *Editaton TEI*. Zenodo. <https://doi.org/10.5281/zenodo.6657093>

5. Workshop “Text encoding Initiative, lenguaje XML y marcado semántico de textos,” at Universitat Catòlica de València, June 29-30, 2022.

CURSO DE HUMANIDADES DIGITALES

Text Encoding Initiative, lenguaje XML y marcado semántico de textos

Plazas limitadas · Inscripción gratuita
Certificado de aprovechamiento de 15 horas

Universidad Católica de Valencia
San Vicente Mártir

Miércoles y jueves
29 y 30 Junio 2022

Sede San Carlos Borromeo
Aula 5002 (informática)
C/ Quevedo, 2. 46001 Valencia

Más información e inscripción en:
<https://bit.ly/3MyAj56>

Objetivos

La edición filológica digital es una de las líneas de investigación más sólidas del campo de las Humanidades Digitales. Mediante el uso de lenguajes de marcado la información textual se codifica para luego recuperarse a través de búsquedas y reusarse en múltiples formatos y contextos de lectura (visualizaciones, grafos, mapas). En este curso reflexionaremos sobre las posibilidades de trabajo filológico y editorial con textos digitalizados y aprenderemos a codificar textos digitales, utilizando el lenguaje de marcado XML-TEI, desarrollado por la Text Encoding Initiative para el intercambio de información textual en el campo de las Humanidades y las Ciencias Sociales. Veremos algunos casos de uso y trabajaremos en modalidad taller sobre la codificación y visualización de algunos ejemplos concretos.

Destinatarios

Docentes, doctorandos y alumnado en general con interés en las Humanidades Digitales.

Metodología

Sesiones teóricas con parte práctica y evaluación final.

Programa

MIÉRCOLES 29

Mañana:

10:00. Humanidades Digitales: Introducción y tipos de proyectos
11:00. Los datos en humanidades ¿cuáles y para qué?
12:00. Pausa café
12:30. La codificación semántica de los textos: La Text Encoding Initiative (TEI)
13:30. Pausa

Tarde:

16:00. El lenguaje XML y las Guías de la TEI
17:00. Modelado de los textos y posibilidades de codificación
18:00. Pausa café
18:30. Manos a la obra: Codificación de un texto modelo
20:00. Fin de la jornada

JUEVES 30

Mañana:

10:00. ¿Qué son y cuál es la finalidad de los "esquema"?
11:00. La función de la ODD
12:00. Pausa café
12:30. Manos a la obra: Personalización de un esquema y una ODD
13:30. Pausa

Tarde:

16:00. ¿Y después de la codificación? Escenarios de transformación
17:30. Pausa café
18:00. Manos a la obra: Transformaciones de XML a HTML
19:30. Balance del curso y turno abierto de preguntas
20:00. Fin de la jornada

Docentes



Susanna Allés Torrent
Associate Professor en la University of Miami (FL). Doctora por la Universidad de Barcelona y diplomada en Humanidades digitales (HD) por la École Nationale des Chartes (Paris) (2012). Lectora en HD en Columbia University (2014-2016), investigadora postdoctoral en la Institució Milà i Fontanals-CSIC y profesora adjunta en la Universitat de Barcelona (2012-2014).



Gimena Del Rio Riande
Investigadora del Instituto de Investigaciones Bibliográficas y Crítica Textual, IIBICRIT-CONICET. Doctora en Filología Románica (premio extraordinario) y Magister en Estudios Literarios por la Universidad Complutense de Madrid. Experta en Investigación y Recuperación del Patrimonio Literario por la Universidad Autónoma de Madrid, y Licenciada y Profesora en Letras por la Universidad de Buenos Aires.

6. Long Paper at TEI Conference and Members' Meeting "Text as Data," Newcastle University (UK), October 25-27, 2022. <https://conferences.ncl.ac.uk/tei2022/programme/>

Where is the Spanish in the TEI?: Insights on a Bilingual Community Survey

Susanna Allés-Torrent (UM)
Gimena del Rio Riande (CONICET)

Who can best define the interests and needs of a community? The members of the community itself.

“Communicating the Text Encoding Initiative to a Multilingual User Community” is a research project financed by the A. Mellon Foundation in which scholars from North and South America are generating linguistic, cultural and didactic situated educational materials to improve XML-TEI encoding, editing and publication of Spanish texts. As it is well known, the TEI has as its primary goal the creation of a model that could semantically describe any text within any cultural heritage, any community, and any language. This goal is only feasible with a diverse and representative group of users that implement the guidelines and the encoding and publishing methodologies, and actively engage in discussions, events and activities. However, when we produce training materials, we often focus on the specific skills, capacities and tools we are trying to teach to individuals. And yet, identifying the community through shared narratives or culture is also crucial.

As part of the project activities, we prepared a bilingual survey (Spanish-English) aimed at inquiring t who uses or has used XML-TEI practices, and where and how they have been applied to Spanish humanistic texts. Bearing in mind that many digital scholarly edition projects of Spanish texts are carried out in Spanish-speaking and Anglophone institutions, we did not focus on a geographical survey, but on the use of XML at a global level. The survey ran between February and April 2022. It is an anonymous survey and consists of 22 questions. It received 104 responses, 77 in Spanish and 28 in English.

Some of the data that we will discuss in this short presentation aims at illustrating the significant differences regarding the organization of projects, collaboration, financing and use of TEI in master's and doctoral research. In broad terms, the survey allowed us to better understand not only the Spanish-speaking community that uses XML-TEI, but also to think of strategies that can contribute with more inclusive practices for scholars from less represented countries and in less favorable contexts inside the global TEI community. Last but not least, we believe the survey will be useful for designing actions that can support a wider range of modes of interaction and collaboration inside the global TEI community.

7. Workshop and presentation at V Conference of the AAHD (Asociación Argentina de Humanidades Digitales) “Humanidades Digitales: Miradas desde el sur,” General Roca-Füskü Menuko, Provincia de Río Negro, Argentina, November 17-18, 2022.

7.1. Workshop “Edición digital de textos con XML-TEI”

Instructors: Gimena del Rio Riande & Susanna Allés Torrent

Abstract: La edición digital es una de las líneas de investigación más sólidas del campo de las Humanidades Digitales. Mediante el uso de lenguajes de marcado como el XML, el HTML y CSS, la información textual se codifica para luego recuperarse a través de búsquedas y reusarse en múltiples formatos y contextos de lectura (visualizaciones, grafos, mapas). En este taller realizaremos una breve introducción a las posibilidades de trabajo filológico y editorial con textos digitalizados y aprenderemos a codificar textos digitales utilizando en lenguaje de marcado XML-TEI, desarrollado por la Text Encoding Initiative para el intercambio de información textual en el campo de las Humanidades y las Ciencias Sociales. Revisaremos algunos casos de uso, los módulos principales de la TEI y codificaremos un texto breve. Usaremos tutoriales y materiales del proyecto TTHub, financiado por la Andrew Mellon Foundation: <https://tthub.io/>.

7.2. Long paper “Codificación de textos y edición digital, ¿Por qué y para qué? Una perspectiva desde el mundo hispanico”.

Susanna Allés-Torrent

???

La codificación de textos ha sido, desde los albores de las humanidades digitales, una de las líneas de trabajo y teorización de esta disciplina. La transformación de los textos en datos informáticos ha constituido una de las cuestiones centrales no sólo para aquellos procedentes del campo de los estudios literarios sino también de todos los estudios humanísticos en general. Además de esta mediación, otro aspecto sobre el que se ha trabajado a nivel internacional ha sido el traer la dimensión semántica a esos datos textuales. Así, por ejemplo, la Text Encoding Initiative ha sido una de las comunidades más activas y que, desde los años ochenta, ha venido

proponiendo un estándar para la codificación de cualquier tipo de textos. La TEI, empleada a nivel internacional y concebida para cualquier lengua, ha tenido un cierto uso entre los proyectos llevados a cabo en el mundo hispánico o relacionados de alguna manera con el español o lenguas afines. Desde el proyecto TTHUB, Hub de Tecnologías del Texto, proponemos una presentación que aborda diferentes cuestiones. En primer lugar, queremos insistir en la relevancia y utilidad de la codificación textual y de la comunidad TEI en particular, con algunos ejemplos específicos. En segundo lugar, ofreceremos algunos de los resultados obtenidos a través de una encuesta llevada a cabo entre los meses de enero y abril de 2022 sobre los usos de TEI en español. En tercer lugar, presentaremos a la comunidad argentina alguno de los nuevos recursos que hemos integrado en nuestra plataforma TTHub con el fin de facilitar y difundir el uso de la TEI para proyectos en español. Estas actividades forman parte del proyecto “Communicating the Text Encoding Initiative to a Multilingual User Community” financiado por la Fundación A. Mellon que tiene como objetivo comprender mejor las necesidades de la comunidad TEI hispanohablante. Por una cuestión de tiempo, la presentación estará a cargo de las investigadoras responsables del proyecto Susanna Allés Torrent y Gimena del Rio Riande.

8. JTEI Special issue

We have received 4 contributions... “Pasado, presente y futuro de la Text Encoding Initiative en español” *Journal of the Text Encoding Initiative*. Special Issue in Spanish

Authors and tentative titles:

1. **Susanna Allés-Torrent & Gimena del Rio Riande**. Introducción.
2. **Antonio Rojas Castro**. Cómo crear ediciones digitales académicas entre La Habana y Berlín. El caso de Proyecto Humboldt Digital.
3. **Ernesto Priani**. Introduciendo filósofos a TEI (en español y a distancia).
4. **Marta López Izquierdo**. Modelado de datos en TEI para la anotación lingüística y pragmática de textos epistolares.
5. **Leyre Martín**. El uso de TEI en la edición y publicación de corpus documentales hispánicos.
6. **José Calvo Tello & Nanette Rissler-Pipka**. ¿Qué hacer con textos que no pueden publicar? Datos derivados, criterios FAIR y TEI en algunos corpus en español.

7. **Gabriel Calarco.** La écfrasis en el *Libro de Alexandre*, un proyecto de edición digital para el estudio de la poesía clerical castellana del siglo XIII con herramientas de minimal computing
8. **Gimena del Rio Riande.**