

**LA RECOLECCIÓN DE DATOS
COMO LABORATORIO EPISTEMOLÓGICO.
ALGUNAS REFLEXIONES ACERCA DEL ENTORNO VIRTUAL
DE INVESTIGACIÓN *REVISTAS CULTURALES 2.0***

DATA COLLECTION AS AN EPISTEMOLOGICAL LABORATORY.
SOME REFLECTIONS ON THE VIRTUAL RESEARCH
ENVIRONMENT *REVISTAS CULTURALES 2.0*

Hanno EHRLICHER
Universidad de Tübingen
hanno.ehrlicher@uni-tuebingen.de

Jörg LEHMANN
Universidad de Tübingen
joerg.lehmann@uni-tuebingen.de

Resumen: Sobre la base de las experiencias ganadas con el entorno virtual de investigación *Revistas Culturales 2.0* queremos reflexionar sobre los retos epistemológicos que supone la recolección de datos. Mientras que los “datos” de la literatura solían considerarse en la investigación filológica como algo previo al proceso epistemológico, en las Humanidades Digitales forman parte integral e incluso pueden ser el resultado de la investigación. Es necesario, pues, empezar a valorar esta importancia epistémica tanto para una mayor aceptación de los métodos de las Humanidades Digitales en las disciplinas filológicas como para mejorar la colaboración entre archivos y bibliotecas, y la investigación en el futuro.

Palabras claves: *Revistas Culturales 2.0*. Entorno virtual de investigación. Revistas culturales. Colección de datos. Epistemología.

Abstract: Based on our experiences gained with the virtual research environment *Revistas culturales 2.0*, we want to reflect on the epistemological challenges of data collection. While in traditional literary criticism “data” used to be considered as something prior to the epistemological process, in Digital Humanities they are an integral part and may even be the result of research. It is necessary to begin to value this epistemic importance both for greater acceptance of the methods of the Digital Humanities in the philological disciplines and to improve collaboration between archives and libraries and research in the future.

Key Words: *Revistas Culturales 2.0*. Virtual research environment. Cultural magazines. Data collection. Epistemology.

1. EL ENTORNO DE INVESTIGACIÓN *REVISTAS CULTURALES 2.0* Y SU POTENCIAL COMO LABORATORIO EPISTEMOLÓGICO PARA LA INVESTIGACIÓN DE REVISTAS CULTURALES

El entorno virtual de investigación *Revistas Culturales 2.0* existe desde 2014 (Ehrlicher, 2016)¹. En él, una red internacional de investigación explora revistas culturales digitalizadas en español. Las copias digitales las proporciona el Instituto Iberoamericano del Patrimonio Cultural Prusiano (IAI, por sus siglas en alemán) como parte de un acuerdo de cooperación y se importan al entorno de investigación a través de una interfaz. En la actualidad ya se puede trabajar con más de un centenar de revistas provenientes de los diversos contextos culturales de América Latina. Y el corpus de las revistas irá creciendo continuamente a medida que aumente el conjunto de las revistas digitalizadas del IAI y, en un futuro, se podría ampliar también sistemáticamente a través de cooperaciones con otras bibliotecas relevantes del mundo hispanohablante. Un ejemplo de ello sería la Biblioteca Nacional de España, en Madrid, que en las últimas décadas ha sido especialmente activa en la digitalización de sus fondos hemerográficos². Al ofrecer la posibilidad de anotar revistas

¹ <https://www.revistas-culturales.de/es> [30/06/2019].

² La Hemeroteca Digital de la BNE ofrece hoy en día más de 2.000 revistas y periódicos, la mayoría

históricas digitalizadas, en nuestro entorno perseguimos el objetivo de una colaboración abierta distribuida (*crowdsourcing*) que consiste en explorar las redes históricas de las revistas culturales en español puesto que estas se pueden considerar transmisoras de la modernidad literaria y sociocultural. Así, el portal desempeña dos funciones: en primer lugar, supera la lógica nacional de las bibliotecas y los archivos que ha llevado también a una disgregación de la investigación especializada a las respectivas dinámicas nacionales sin poder llegar a una comparación transnacional. En el espacio cultural hispanohablante hay que constatar una falta de visiones comparativas no solamente en lo referente al intercambio transatlántico entre España y los países latinoamericanos, sino también en cuanto al intercambio inter- y transnacional dentro del espacio cultural americano. Se trata, por supuesto, de un problema que viene de lejos y que llevó a que se hablara de los “Estados Des-unidos de la América del Sur” —en contraste con los EE.UU.—, apenas unas décadas después de que lograran la independencia la mayoría de sus países³. Sin embargo, nuestro entorno de investigación no solo les proporciona a los investigadores situados en diferentes partes del mundo hispánico el acceso a las copias digitales y a los metadatos correspondientes, sino que también les permite realizar planteamientos individuales y recopilar datos relevantes para sus propios proyectos. Esto último es la segunda plusvalía crucial de este entorno de investigación. Además de los archivos de imagen de las revistas culturales digitalizadas, los metadatos proporcionados por el IAI también se importan en el estándar METS/MODS basado en XML (*Metadata Encoding & Transmission Standard / Metadata Object Description Schema*). Mediante formularios de anotación preestructurados, el portal también le permite al colectivo de los investigadores ampliar estos metadatos sistemáticamente (figuras 1 y 2).

de ellos son de libre acceso: http://www.bne.es/es/Catalogos/HemerotecaDigital/docs/tabla_listadocompleto.pdf [30/06/2019].

³ La expresión “Estados Des-unidos” la utilizó, por ejemplo, el intelectual chileno Francisco Bilbao en el *post dictum* a un discurso pronunciado en París, en 1856. Bajo el título de “Iniciativa de América. Idea de un Congreso Federal de las Repúblicas”, el autor retoma la vieja idea bolivariana de una Unión Federal de Estados Latinoamericanos. En el *post dictum* explica sus objetivos de la siguiente manera: “La idea de la Confederación de la América del Sur, propuesta un día por Bolívar, intentada después por un Congreso de plenipotenciarios de algunas de las Repúblicas, y reunido en Lima, no ha producido los resultados que debían esperarse. Los Estados han permanecido *Des-Unidos*” (Bilbao, 1866: 285).

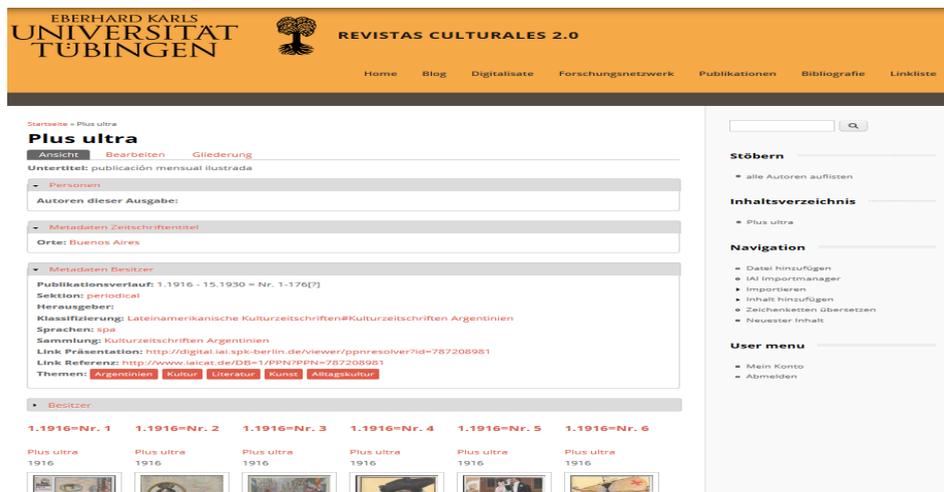


Figura 1. Captura de pantalla del entorno de investigación virtual *Rev*



Figura 2. Captura de pantalla del entorno de investigación virtual *Revistas Culturales 2.0*; una página de la revista cultural *Plus Ultra*.

Como bien muestran las dos capturas de pantalla, a los investigadores se les presentan primero los metadatos recogidos por el IAI. En el formulario ofrecido al usuario en el entorno de investigación, los investigadores pueden recopilar más datos. Existe la posibilidad de

seleccionar con un clic categorías preestablecidas o rellenar determinados campos de datos (por ejemplo, los títulos de los artículos). Pero también hay la posibilidad de añadir nuevas categorías (por ejemplo, los géneros de texto) de forma acumulativa o de realizar anotaciones libres siguiendo los objetivos personales de la investigación propia. Después de haber trabajado con una revista, todos los datos recopilados se pueden guardar en el banco de datos del entorno y exportarlos en formato CSV.

Esto ya muestra una diferencia fundamental con respecto a los métodos clásicos de la investigación de la literatura. Esta suele tomar un solo texto o un grupo de textos como base de la investigación considerando que sus “datos” son anteriores al proceso epistémico. Tanto el texto en sí como las indicaciones acerca del autor, la editorial, el año de publicación, etc., forman en su conjunto los datos de los estudios literarios. Son considerados como previos al proceso epistemológico y como ya dados, que es, precisamente, lo que significa la palabra *datum*. En realidad, hoy en día, muchos análisis de *big data*⁴ proceden de esta manera y esta tendencia se ve también en las muy discutidas propuestas de una lectura distante (*distant reading*) de Franco Moretti y sus colaboradores. Suelen partir de un archivo literario cuantitativamente ampliado a escala global y dejan la “lectura” de los datos literarios a cargo de las máquinas, pero heredan las categorizaciones y con ellas la estructuración de los datos de las filologías nacionales cuyos presupuestos siguen aceptando como si se tratara de resultados empíricos incuestionables⁵. Por el contrario, en el entorno de investigación virtual *Revistas Culturales 2.0* se deben recoger primero activamente los mismos datos que constituyen la base de la evaluación posterior. Esto muestra que los datos pueden ser vistos tanto como el inicio de un proceso de investigación como su propio resultado. Este último es

⁴El hecho de que los *big data* son, casi siempre, anteriores al proceso epistémico quedó demostrado en el proyecto K-PLEX (Lehmann, Huber y Stodulka, 2018: 75).

⁵Franco Moretti es un científico bastante libre en cuanto a su metodología y está dispuesto a falsificar sus propias presuposiciones, por lo tanto, hay que ser preciso aquí. Al principio de su discurso sobre el *distant reading*, Moretti parte de una suerte de “cosmic and inevitable division of labour” (2000: 66), en la que el comparatista opera desde la distancia y emplea los resultados que ha adquirido de la investigación de las filologías nacionales, por lo que acredita estas investigaciones ciegamente como si fueran datos empíricos y no interpretaciones. Sin embargo, Moretti ha destacado recientemente (2013) el potencial de su método de producir un extrañamiento epistémico y subraya que los procesos de operatividad pueden poner en duda las categorías inicialmente presupuestas y alterar los resultados.

el caso si son el resultado de una investigación en el marco de un proceso epistemológico, lo cual es evidente en las Ciencias Sociales. En ellas no solo se procesan los datos económicos y socioestructurales –tales como los que proporcionan las estadísticas oficiales de DeStatis, Eurostat y de la ONU–, sino que una parte significativa de los datos tiene que ser recopilada por parte de los investigadores, especialmente a través de cuestionarios.

Sin embargo, tal enfoque –igual que la sofisticada metodología que surge de él– no es tan evidente en las Humanidades. Esta falta de una tradición metodológica en las Humanidades explica también, en parte, por qué la comunidad internacional de investigadores ha tardado bastante en ir aceptando las ofertas de recopilación de datos como las que se ofrecen en nuestro entorno de investigación. Los filólogos especializados en literatura, por ejemplo, tendrán que adquirir primero los métodos de clasificación y evaluación. El reto epistemológico para las Humanidades consiste en que la clasificación formalizada reduce la flexibilidad acostumbrada, pues las clasificaciones que tienen que ser formalmente estables obligan a una fijación que no es propia de las interpretaciones; al cambiar el marco interpretativo, las clasificaciones tienen que ser adaptadas *a posteriori*, lo que implica una inversión de tiempo enorme. Además, este desafío epistemológico va acompañado del obstáculo sociopragmático de una falta de gratificación del esfuerzo involucrado, ya que las clasificaciones, tan importantes en la recopilación de datos, no solo requieren mucho tiempo, sino también muchos recursos. En las filologías, independientemente si se siguen los procedimientos clásicos de interpretación hermenéutica o métodos cualitativos post-hermenéuticos, el valor cognitivo invertido en la colección de datos normalmente no será valorado como productivo, sino que desaparecerá completamente detrás de las conclusiones interpretativas que se justifican argumentativamente como tesis. Basta con tener en cuenta las semanas y los meses invertidos para la recopilación de datos en nuestro proyecto para evaluar el enorme desequilibrio entre la importancia práctica de la recopilación de datos y el riesgo de una falta de reconocimiento académico. Los investigadores del ámbito tradicional de las filologías que no están implicados en nuestro proyecto tienen por lo tanto bastantes razones pragmáticas para no querer exponerse a este riesgo sin más. En las Humanidades todavía no hay sensibilidad suficiente o normas de *best practice* para valorar la recopilación de datos como parte sustancial y costosa del proceso epistémico. Pero solo empezará a formar parte de

los métodos habituales de una disciplina académica siempre y cuando sea tomada en consideración en los sistemas de evaluación y gratificación esta disciplina.

Para contribuir al objetivo de aumentar la valorización de la recolección de datos en las Humanidades ofrecemos nuestra reflexión metodológica que se basará aquí en ejemplos concretos; precisamente porque el carácter eminentemente práctico de la recopilación de datos es de naturaleza praxiológica. En realidad, con las consideraciones siguientes enfocaremos un nivel epistémico intermedio, que no fue tenido en cuenta, por ejemplo, en las interesantes y recientes reflexiones de Peer Trilcke y Frank Fischer sobre la “praxiología de los estudios literarios digitales y sus ‘cosas’ epistémicas” (2018). Este nivel epistémico intermedio es, sin embargo, decisivo en el marco de nuestra práctica de investigación. En general, también debería ser de gran importancia para una mediación exitosa o, al menos mejorada, entre las “cosas epistémicas” ligadas al medio textual del que tratan los estudios literarios clásicos y las “cosas” basadas en códigos que son modeladas y analizadas por los nuevos estudios literarios digitales. Este problema general se explicará ahora con más detalle mediante ejemplos concretos tomados del proyecto „Procesos de modernización literaria y creación de redes transnacionales en el medio de la revista cultural“, el cual está estrechamente relacionado con el entorno de investigación virtual *Revistas Culturales 2.0*⁶.

2. LA RECOLECCIÓN DE METADATOS COMO DESAFÍO EPISTÉMICO

En este proyecto de investigación, se examina un corpus textual de casi 60 revistas digitalizadas del ámbito cultural de habla hispana. Metodológicamente el centro de interés son las redes transnacionales de actores sociales que se pueden detectar en este corpus así como la red de las formas textuales (géneros y tipos de texto) que es prueba de una transferencia cultural de las literaturas. La extensión intercultural y transatlántica de este corpus de revistas puede verse ilustrada en la

⁶Proyecto financiado por la Fundación Alemana de Investigación Científica (DFG) con el número 327964298. Véase el breve resumen en el sistema de información correspondiente <http://gepris.dfg.de/gepris/projekt/327964298> [30/06/2019].

contribuciones de los autores a larga escala temporal y espacial. Entonces se hace evidente la importancia de transferencias culturales transcontinentales que están en el centro del interés de nuestro proyecto de investigación. Como prueba, ya puede ser ilustrada con la visualización de una pequeña sección de nuestro corpus (figura 4).

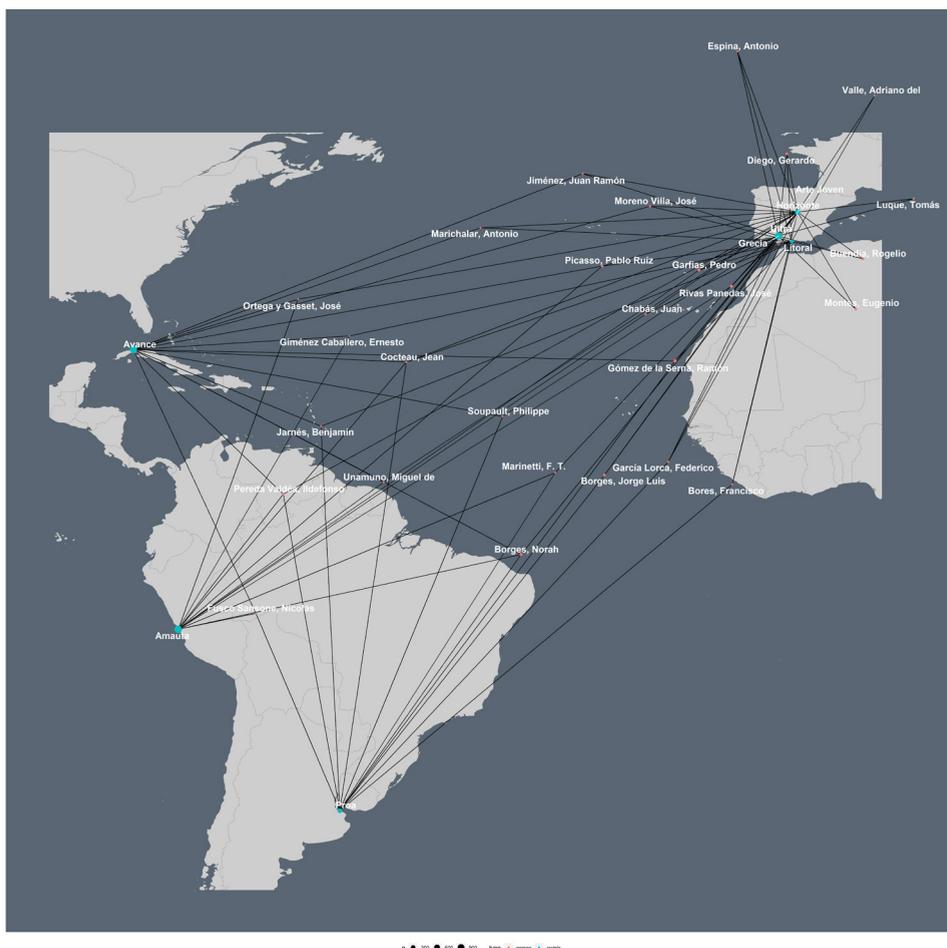


Figura 4. Visualización de todos aquellos colaboradores que han realizado al menos 2 contribuciones en cada una las siguientes ocho revistas culturales: *Arte Joven* (Barcelona), *Grecia* (Sevilla) *Horizonte* y *Ultra* (Madrid), *Litoral* (Málaga), *Proa* (Buenos Aires), *Amauta* (Lima) y *Avance* (La Habana).

Como en otros proyectos de las Humanidades Digitales, nuestro enfoque se caracteriza por dejar atrás el análisis de un solo objeto para centrarnos en el análisis de estructuras o “patrones” comparativos en un *corpus* tan grande que ya es imposible leerlo tan solo de modo hermenéutico. Con el cambio de la escala del objeto no solamente cambia nuestro modo de entenderlo sino también el mismo objeto de estudio:

One thing for sure: digitization has completely changed the literary archive. People like me used to work on a few hundred nineteenth-century novels; today, we work on thousands of them; tomorrow, hundreds of thousands. This has had a major effect on literary history, obviously enough, but also on critical methodology; because, when we work on 200.000 novels instead of 200, we are not doing the same thing. .,000 times bigger; we are doing a different thing. The new scale changes our relationship to our object, and in fact it changes the object itself (Moretti, 2017: 1).

Sin embargo, a diferencia de Moretti, en nuestro proyecto no queremos tratar tan solo el nuevo “objeto” de estudio que emerge de la posibilidad de una lectura distante, sino mediar entre este nuevo objeto, el texto tradicional y los métodos de entenderlo, en un enfoque mixto que combina lo cuantitativo y lo cualitativo. Visualizaciones como la que muestra la figura 4 nos servirán en el proyecto para descubrir actores cuyo rol dentro de la red de contribuciones se tiene que investigar después con lecturas detalladas de los textos respectivos. No obstante, la problemática epistémica inherente a la recolección de datos es, lógicamente, anterior a la posibilidad de establecer y reconocer patrones cuantitativos.

Como ya se ha mencionado, es precisamente este proceso preparativo —que requiere tanto tiempo ya que se lleva a cabo manualmente y, a primera vista, puede parecer un asunto meramente mecánico— el que queremos dar visibilidad y reconocimiento aquí. Un primer examen detenido de la recopilación de datos en el entorno de investigación virtual muestra cuán problemático es, de hecho, hablar de “dato” en el sentido de algo “dado”. Aunque el IAI proporcione los metadatos sobre los autores, el lugar de publicación o los géneros utilizados, sin embargo, las contribuciones anónimas, las contribuciones de autoría colectiva o los seudónimos —que, debido a la censura, eran la

regla más que la excepción en las revistas del modernismo⁷— demuestran que detrás de estas clasificaciones aparentemente claras se pueden ocultar casos problemáticos cuya resolución requiere enormemente tiempo de investigación minuciosa. También la recopilación de los datos básicos de las publicaciones periódicas resulta problemática en algunos casos y no es en absoluto una tarea baladí. Porque, más allá de su función de receptáculo de textos de diferente procedencia cultural, algunas revistas culturales como *Alfar* o *Creación/Création* eran, a su vez, objetos móviles que por diversas razones fueron distribuidos por sus editores en diferentes lugares, por lo que se plantea la cuestión de si deben ser evaluados como una sola revista o como varias. Como ejemplo cabe mencionar la revista *Alfar*, publicada por Julio J. Casal. Nacido en Montevideo de una pareja hispano-uruguaya, su carrera diplomática le llevó primero a España, donde publicó la revista en La Coruña como continuación del *Boletín de la Casa América-Galicia* bajo el nuevo título de *Alfar* desde el número 33 (octubre de 1923) hasta el número 60 (1926). Tras unos años de interrupción, siguió con el proyecto en Uruguay desde 1929 hasta 1954. En ausencia de Julio J. Casal, un antiguo empleado editó tres números más en La Coruña, por lo que a veces había dos ramales que se hacían la competencia entre sí. Esta publicación —de gran importancia no sólo para la representación de las literaturas no castellanas de la Península Ibérica de su época, sino también para el intercambio cultural transatlántico— ha sido tratada de forma diferente por la investigación. La única monografía detallada publicada hasta la fecha en España la trata como una revista exclusivamente en el contexto nacional y, por lo tanto, prescinde de su historia posterior en Uruguay (Molina, 1984). Mientras tanto, la plataforma central para la presentación de publicaciones periódicas digitales en Uruguay evalúa la revista, claro está, como “uruguaya”⁸.

Si, en este caso, la movilidad de la revista se explica únicamente por las circunstancias biográficas del editor, entonces la movilidad

⁷ Para la fase del modernismo en los diferentes contextos nacionales del ámbito cultural hispanohablante, nos hemos servido de diccionarios como el de Figarola-Candea (1922) o de Scarone (1942) para desentrañar los seudónimos más frecuentes de los autores.

⁸ <http://anaforas.fic.edu.uy/jspui/handle/123456789/5562> [30/06/2019]. Con todo, tiene la gran ventaja de ofrecer copias de ambas fases de la revista, mientras que el facsimilar realizado por la editorial Nos en 1983 abarca únicamente la producción peninsular.

intercultural de *Creación / Création* –una revista publicada inicialmente en varios idiomas por Huidobro– también es programática en términos de contenido. En su visión de una comunidad utópica de artistas, el autor chileno vanguardista parece haber asumido como algo natural que las fronteras lingüísticas y nacionales ya no juegan un papel internacional de las artes, que él quería interconectar en su foro. En la clasificación bibliográfica de una revista de este tipo o bien se ignora conscientemente ese carácter intercultural o bien se hace difuso debido a las normativas pragmáticas. La Bibliothèque Nationale de Francia, por ejemplo, separa el primer número multilingüe de la revista publicada en España de las demás ediciones en francés y lo describe como una revista secundaria exótica (Figuras 5 y 6). Por otra parte, el IAI, que digitalizó la revista para nuestro proyecto de investigación, ofrece metadatos que son obviamente inconsistentes porque la revista es contraria a los principios de clasificación inequívocos (figura 7).

The screenshot shows the BnF Catalogue général search results for the journal 'Creación'. The interface is in French. At the top, there is a search bar with the text 'Tout' and a magnifying glass icon. To the right of the search bar, there are navigation tabs: 'Recherche avancée', 'AUTEURS A-Z', 'SUJETS A-Z', 'PÉRIODIQUES', 'COTE', and 'NOTICES'. Below the search bar, there is a breadcrumb trail: 'Accueil > Recherche avancée > Liste de notices'. The main content area shows '1 Notice bibliographique'. On the left, there is a sidebar with 'VOTRE RECHERCHE' and 'AFFINER' sections. The 'VOTRE RECHERCHE' section shows the search criteria: 'Recherche avancée : Dans toute la notice : Creación Revista internacional de arte / Tous les mots'. The 'AFFINER' section shows filters for 'Nature de document', 'Périodiques (1)', 'Texte imprimé et livre numérique (1)', 'Détails texte imprimé ou électronique', 'Périodiques (1)', and 'Pays', with 'inconnu (1)' selected. The main results area shows one result: '*Creación. Revista internacional de arte Périodiques et collections'. The result is displayed in a table with columns for 'Tri par: Défaut', '1 sur 1', and '10 résultats/page'. The result title is '*Creación. Revista internacional de arte' and the description is 'Périodiques et collections'. The result is displayed in a table with columns for 'Tri par: Défaut', '1 sur 1', and '10 résultats/page'.

Figura 5. La revista *Creación* en el catálogo de la *Bibliothèque Nationale de France* (BnF).

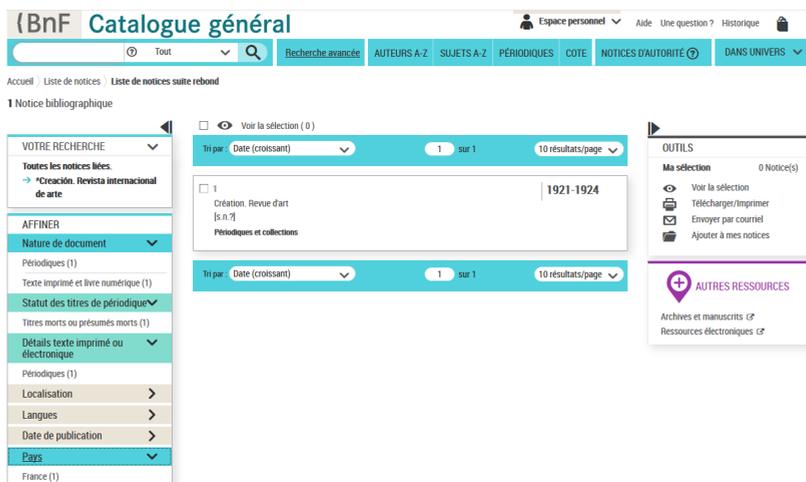


Figura 6. La revista *Création* en el catálogo de la *Bibliothèque Nationale de France* (BNF)

```

-<smets:mets:OBJID="" xsi:schemaLocation="http://www.loc.gov/mods/v3 http://www.loc.gov/standards/mods/v3/mods-3-5.xsd
v2.0.xsd http://www.loc.gov/standards/mix http://www.loc.gov/standards/mix/mix.xsd">
-<smets:metsHdr:CREATEDATE="2018-02-06T14:34:29Z">
-<smets:agent:OTHERTYPE="SOFTWARE" ROLE="CREATOR" TYPE="OTHER">
-<smets:name>
Goobi- ugh-3.0-ugh-2.0.0-29-g3b6fef1 - 21-December-2016
</smets:name>
<smets:note>Goobi</smets:note>
</smets:agent>
</smets:metsHdr>
-<smets:dmdSec ID="DMDLOG_0000">
-<smets:mdWrap MDTYPE="MODS">
-<smets:xmlData>
-<mods:mods>
-<mods:originInfo>
-<mods:place>
<smets:placeTerm type="text">Madrid, Paris</mods:placeTerm>
</mods:place>
<smets:dateOther>Número 1 (abril 1921)-février 1924</mods:dateOther>
</mods:originInfo>
<smets:classification authority="ivdc">Varia</mods:classification>
</mods:recordInfo>
<smets:recordIdentifier source="gbv-ppn">1012316823</mods:recordIdentifier>
</mods:recordInfo>
<smets:identifier type="url">
http://resolver.iai.spk-berlin.de/IAI0000695800000000
</mods:identifier>
<smets:identifier type="zdb">2918814-3</mods:identifier>
</mods:titleInfo>
<smets:title>Créación</mods:title>
</mods:titleInfo>
<smets:titleInfo type="alternative">
<smets:title>Création <span></smets:title>
</mods:titleInfo>
</mods:subject authority="gnd">
<smets:geographic>Spanien</mods:geographic>
</mods:subject>
<smets:subject authority="gnd">
<smets:geographic>Frankreich</mods:geographic>
</mods:subject>
<smets:subject authority="gnd">
<smets:topic>Kunst</mods:topic>

```

Figura 7. La revista *Creación* registrada en los metadatos recogido por el Instituto Iberoamericano (IAI).

Incluso en este primer nivel de la concepción del objeto de estudio, resulta claro que la multidimensionalidad del objeto socava la consistencia de un sistema de clasificación preestablecido y configurado para una larga duración. En este caso concreto, la ambigüedad de la clasificación geográfica inherente a la revista ha sido tratada de manera diferente por el personal bibliotecario ya que, por una parte, en la categoría de clasificación “place” se ha decidido por una doble entrada “Madrid-París”, mientras que bajo “subject” se ha codificado el lugar de procedencia de la revista (“subject, geographic”) dos veces (“Frankreich”, es decir, Francia, y “Spanien”, España). Sin embargo, el hecho de que el idioma de la descripción de las categorías (inglés) choque con el idioma de la categorización (alemán) ya no se debe al objeto, sino que se explica por las reglas internas, pero históricamente heredadas, de la catalogación del IAI. Aquí cabría preguntarse hasta qué punto es útil ofrecer hoy en día aún metadatos en alemán a un círculo internacional de usuarios que maneja con seguridad el español y el inglés, pero con mucha menor probabilidad el alemán.

Sin embargo, el problema epistémico de la categorización también se hace evidente cuando nos alejamos del nivel de metadatos prescrito por la biblioteca y miramos más de cerca los datos capturados de forma independiente por los investigadores, como son los de los autores de las contribuciones. “Datos” aparentemente tan inequívocos como el sexo o la nacionalidad de un autor son muy cuestionables en nuestro mundo moderno, marcado por las historias de migración y exilio ya desde el siglo XX⁹, y marcado, al comienzo del siglo XXI también por

⁹ Resaltamos aquí tan solo un autor especialmente conocido de nuestro *corpus* literario, Max Aub Mohrenwitz (1903-1972). En su caso, la problemática de identificar claramente una “nacionalidad” se puede ver ya al intentar retomar las clasificaciones propuestas por el *Virtual International Authority File* (<https://viaf.org/> [30/06/2019]) que hemos utilizado como referencia estándar por su uso internacional. A Max Aub se le clasifica ahí de autor tanto “francés” como “mexicano” y —mayoritariamente— “español”, y todas las clasificaciones son acertadas, aunque tan solo parcialmente, lo que deja entrever muy bien la problemática de querer fijar en tan solo una “nacionalidad” una historia personal tan compleja y migratoria como la de Aub. Nacido en Francia, pero de padre alemán y madre francesa, se nacionalizó primero en España y tuvo que exiliarse a México, donde más tarde obtendría la nacionalidad mexicana.

cuestiones de género que han llevado a introducir políticamente categorías como la ‘intersexualidad’ que van más allá de la binaria masculino/femenino. En un mundo de datos que se centra en unidades contables y observables, las categorías “hombre”, “mujer” o “ciudadanía” se dan por sentadas y se consideran como “diferencias naturales”, aunque estén construidas socialmente (Westbrook y Saperstein, 2010). Los términos polivalentes y los conceptos ambiguos no tienen cabida en estos sistemas de clasificación, o a lo sumo en la categoría colectiva de “otros”. Estos ejemplos ya muestran que un enfoque positivista de la realidad postula un lenguaje neutro de observación y una doctrina de ‘datos puros’ que existen independientemente de las teorías, y ambos presupuestos son altamente problemáticos por ‘naturalizar’ las ideologías inherentes en ellos. En contraste con este positivismo de datos, el trabajo en nuestro entorno de la investigación deja claro que toda observación se basa en presupuestos teóricos y que, por lo tanto, no es posible llegar a una evaluación neutral; cada observación está guiada más bien por el interés y cada conocimiento se construye socialmente.

La convención social sobre lo que se considera un “hecho” resulta particularmente evidente si se toma el ejemplo de los géneros textuales. Por supuesto, se pueden encontrar categorías generales como “image”, “review”, “non-fictional prose”, “fictional prose”, etc. a las que se pueden asignar las contribuciones individuales de una revista. Estas clasificaciones, sin embargo, a menudo entran en conflicto con las clasificaciones utilizadas por las propias revistas, que normalmente se encuentran en los índices y que, vistas en conjunto, ofrecen una variedad inconmensurable de clasificaciones de género. De ahí que, de acuerdo con Johanna Drucker (2011), tenga sentido descartar el término “datum” y utilizar en su lugar “captum”, algo “tomado” del enfoque teórico desde el que se considera el material de investigación. Esto abre la posibilidad, por ejemplo, de clasificar los géneros como categorías superpuestas, cuya definición más precisa se debe entender de acuerdo con su proximidad a otros géneros en el contexto de una revista; y todos ellos, en conjunto, constituirían el perfil de esta (Figura 8).

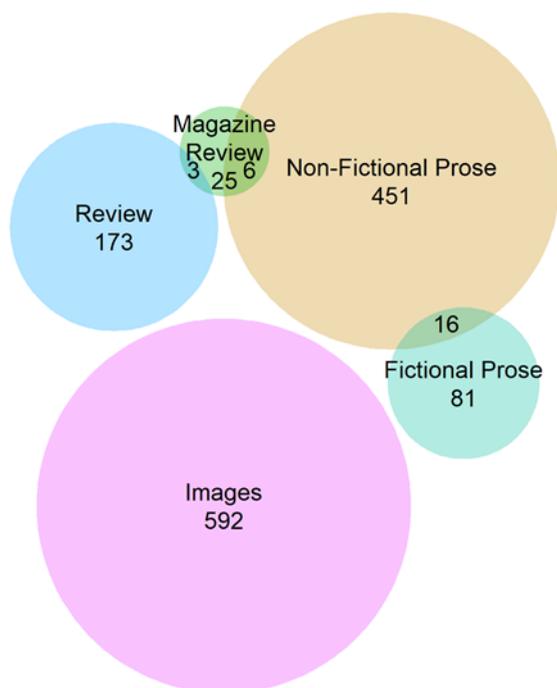


Figura 8. Visualización de géneros que se solapan en la revista cultural Avance.

Visto lo dicho, no es de extrañar que los investigadores que participan en el proyecto de investigación “Procesos de modernización literaria y creación de redes transnacionales en el medio de la revista cultural” hayan ido cambiando y perfeccionando permanentemente, durante los meses de captura de datos, su sistema de clasificación de géneros. Hay un total de tres columnas que describen los géneros utilizados: en la columna “Genre type: comparable” se utilizan abstracciones que sirven para hacer comparaciones entre revistas, como “non-fictional prose”, “image” o “review”. En la columna “Genre type: exact” se utiliza una clasificación más precisa; el término generalizador “non-fictional prose” se concreta aquí, por ejemplo, con “note” (para comentarios). Finalmente, en la tercera columna “Genre type: specific” se especifica la terminología que figura en la segunda columna. La “note” puede, por ejemplo, convertirse

en una “programmatic note”, “literary note” o “autorreflexive note”. La proliferación de estas taxonomías es inevitable, al igual que la discusión de casos problemáticos, como la autoría colectiva de un texto o la combinación de diferentes géneros en un solo texto. Estos ejemplos dejan claro que la normalización de los datos, es decir, el intento de encontrar clasificaciones válidas que eviten redundancias, se convierte en un apretado corsé del que el afán hermenéutico de entender el caso concreto quiere escaparse. Esto resultará evidente para los filólogos especializados en la literatura, ya que están acostumbrados a que los términos utilizados en su campo de investigación se deslicen semánticamente dando lugar a la polisemia y la ambigüedad. Por último, también debe tenerse en cuenta que la estructura de los datos debe obedecer a los requisitos de la evaluación digital para que los datos puedan utilizarse eficazmente (figura 9).

Type: comparable	Type: exact	Type: specific	Colour key
Image	Drawing	Charcoal Drawing	Needs more clarification = educated guess
Non-Fictional Prose	Comment	Editorial mission statement	Check further nationality
Review	Culture Review		No information
Fictional Prose	Short Story		Contributor name given but unreadable
Image	Drawing		Ambiguous assignment/Educated guess
Lyricism	Poem		Pseudonym (possibly)
Image	Drawing	Charcoal Drawing	No information available
Non-Fictional Prose	Comment		Information might be found via other sources
Lyricism	Prose		
Fictional Prose	Short Story		
Image	Drawing	Charcoal Drawing	
Image	Drawing	Charcoal Drawing	
Lyricism	Poem		
Lyricism	Poem		
Non-Fictional Prose	News tidbits		
Image	Drawing		
Non-Fictional Prose	Comment		
Image	Drawing		
Lyricism	Poem		
Fictional Prose	Book fragment		
Image	Drawing		

Figura 9. Ejemplo de la hoja de datos de la revista *Arte Joven* (con comentarios de la investigadora).

Esta creciente diferenciación de un sistema de clasificación pone de manifiesto lo que Ludwik Fleck (1980) ya explicó en su trabajo, ya clásico, sobre la aparición y el desarrollo de un “hecho científico”¹⁰: la percepción del mundo se habitúa, es decir, se convierte en un hábito de una manera muy específica. A este hábito también se debe que, después de algún tiempo, se

¹⁰ Véase al respecto también Lorraine Daston (2008).

olvide que un objeto ha sido declarado hecho científico por convención y aprendizaje, por ejemplo, cuando los patrones recurrentes solo se hacen visibles cuando ya se han examinado miles de páginas de revistas. Los hábitos de percepción adquiridos permiten la observación aparentemente neutral y desinteresada, lo cual es la base de la ciencia. Los investigadores son conscientes de la adquisición de estas habilidades, aunque la acepten como conocimiento tácito. Aprender a ver el mundo como un científico es una cuestión de experiencia acumulada. En el entorno de investigación virtual, la que acumula los datos es la colectividad de los investigadores, no una persona individual. De esa manera, los científicos desarrollan y aprenden modos de percepción como parte de su deformación profesional. El ejemplo de las *Revistas Culturales 2.0* muestra de manera contundente cómo cambia la percepción de estas clasificaciones por parte de los investigadores cuando se visualizan cientos o, incluso, miles de páginas de estas representaciones digitales.

3. EL PESO DE LA INFRAESTRUCTURA: LA INTERACCIÓN ENTRE LA INVESTIGACIÓN Y LA BIBLIOTECONOMÍA EN EL CURSO DE LA DIGITALIZACIÓN DE LOS BIENES CULTURALES

Si se da un paso atrás y se observa la interacción entre el investigador individual, el entorno de investigación y los metadatos proporcionados por las bibliotecas, queda claro que se trata de un proceso de intercambio a gran escala. ¿Son realmente los investigadores los que deciden sobre un cierto tipo de datos y métodos específicos o ya está preformada esta decisión por los objetos existentes y las herramientas y métodos apropiados? (Gray *et alii*, 2018). Incluso antes de que los investigadores individuales recopilen los datos, su percepción de las categorías está preformateada por el origen de estos datos en índices y registros. Esto se refleja en el entusiasmo de las Humanidades Digitales con los índices de sujetos, temas, personas y lugares que se corresponden con técnicas como el *topic modeling* o *named entity recognition*; o también en su preferencia por las visualizaciones como mapas y cronologías (líneas de tiempo). Todos estos ejemplos se refieren a sistemas de ordenación y categorización como los que ya existían en libros

históricos y que son, por lo tanto, previos al conocimiento.

Sin embargo, no es solo el mundo de los libros y la formación de los investigadores en el sistema universitario lo que influencia en su percepción de los procesos epistémicos. Las poderosas infraestructuras entran en juego más bien a través de los metadatos de las bibliotecas –en nuestro caso, los datos recogidos por el IAI–. Las editoriales crean sus libros como datos de texto estructurados en formatos como SGML e introducen los datos descriptivos y los textos de portada a las máquinas de *marketing* como el catálogo de libros en venta de los distribuidores en Alemania (VLB: *Verzeichnis Lieferbarer Bücher*). Las bibliotecas adquieren estos metadatos en formato XML y los procesan en sus bases de datos internas y en los OPAC. La investigación mediante estas herramientas proporciona resultados prefijados basados en estos metadatos, así como en estándares de bibliotecas clásicos e históricamente desarrollados. Si se tiene en cuenta que en el Tercer Reich alemán, en la República de Weimar, en la República Democrática y República Federal de Alemania entre 1899 y 1983 se utilizaban las normas bibliotecarias de las Instrucciones Prusianas (PI, por sus siglas en alemán) y que entre 1976 y 2015 se aplicaron las normas para la catalogación alfabética (RAK, por sus siglas en alemán), y que desde 2015 se ha utilizado el estándar *Ressource Description & Access* (RDA), también hay que tener en cuenta las implicaciones que conllevan dichas normas. Por poner un ejemplo llamativo: el *Dewey Decimal Classification System* (DDC), que todavía se utiliza a nivel internacional, tiene diez categorías para “religión” (DDC 200). Nueve de ellas se utilizan para religiones cristianas, mientras que solo una (DDC 290) puede utilizarse para todas las religiones no cristianas y la comparación de religiones (figura 10).

290 Other & comparative religions	<ul style="list-style-type: none"> • 290 Other & comparative religions • 291 Comparative religion <ul style="list-style-type: none"> ◦ 291.1 Religious mythology, social theology, interreligious relations and attitudes ◦ 291.2 Doctrines ◦ 291.3 Public worship and other practices ◦ 291.4 Religious experience, life, practice ◦ 291.5 Moral theology ◦ 291.6 Leaders and organization ◦ 291.7 Missions, religious training and education ◦ 291.8 Sources ◦ 291.9 Sects and reform movements • 292 Classical (Greek & Roman) religion • 293 Germanic religion • 294 Religions of Indic origin <ul style="list-style-type: none"> ◦ 294.3 Buddhism ◦ 294.4 Jainism ◦ 294.5 Hinduism ◦ 294.6 Sikhism ◦ 294.7 Zoroastrianism (Mazdaism, Parseeism) • 295 Judaism <ul style="list-style-type: none"> ◦ 295.1 Sources ◦ 295.3 Doctrinal, moral, social theology ◦ 295.4 Traditions, rites, public services ◦ 295.6 Leaders, organization, religious education ◦ 295.7 Religious experience, life, practice ◦ 295.8 Sects and movements • 297 Islam & religions originating in it • 298 • 299 Other religions
-----------------------------------	--

Figura 10. La clasificación DDC-290 “Otras religiones y religiones comparadas”.

Una comparación de estos conjuntos de normas revelaría inmediatamente la inconmensurabilidad de estos sistemas de clasificación a lo largo del tiempo. En este caso, sería necesario discutir hasta qué punto estos sistemas de clasificación se corresponden con los patrones de pensamiento contemporáneos de los usuarios. Por último, habría que preguntarse en qué ámbitos existen grandes similitudes entre las categorías de clasificación de las bibliotecas y las clasificaciones de productos asignadas por los editores (como el *Bureau International des Containers BIC*, *Book Industry Subject and Category BISAC*, el estándar de xml *Online Information eXchange ONIX*, por nombrar solo las que se utilizan con mayor frecuencia). El ejemplo de las bibliotecas ilustra que los sistemas de clasificación son difíciles de analizar: cuanto mejor funcionan y cuanto mejor los aceptan los usuarios, menos visibles son. Y cuanto más completos y poderosos son, menos visibles son (Bowker, Leigh Star, 1999). En general, estos sistemas se consideran, por tanto, como “datos” (*datum*) incuestionables e inmutables. Como predijo Jean-François Lyotard ya en 1979, las bases de datos son “la ‘naturaleza’ del hombre postmoderno” (1988: 85).

Los sistemas de clasificación y estandarización forman vínculos entre la organización social de una sociedad, el orden moral y los sistemas de integración técnica. En última instancia, queda claro que la construcción de conocimiento es un proceso cíclico en la interacción entre el investigador individual, el entorno de investigación y los metadatos proporcionados por las bibliotecas: Las convicciones sobre la naturaleza del mundo influyen

en el diseño de los sistemas de clasificación, sobre la base de los cuales se recopilan los datos. La búsqueda en los catálogos de las bibliotecas y la explotación de los metadatos forman a su vez los conceptos de los usuarios y científicos sobre la naturaleza del mundo.

Pero, ¿cómo se relacionan los metadatos disponibles en las grandes infraestructuras con los datos recopilados en el entorno de investigación virtual? A nivel técnico es posible devolver y convertir los metadatos recogidos en nuestro proyecto de investigación –en el que han introducido los conocimientos de la comunidad investigadora especializada en el material–, al formato METS/MODS de las bibliotecas. Sin embargo, nuestra experiencia demuestra que la capacidad institucional por rescatar los metadatos recopilados por investigadores es baja. Por un lado, se debe a la falta de recursos humanos; por otro lado, a razones técnicas, como la necesidad de versionar y la provisión de memorias de datos cada vez mayores. El formato METS/MODS es realmente flexible y abierto para el enriquecimiento de los registros de datos. Sin embargo, las respuestas clasificatorias flexibles y adecuadas al material desarrollado en nuestro pequeño y ágil portal aún están lejos de migrar a los sistemas de clasificación de las bibliotecas. Estos muestran más bien inercia y tradicionalismo, lo que obstaculiza un desarrollo dinámico necesario para cooperaciones con grupos de investigación. Cabe señalar que las infraestructuras de datos pueden seguir actuando como fuerza normativa (Gray *et al.* 2018: 3). Las infraestructuras digitales conforman así el espacio de pensamiento en el que se lleva a cabo la investigación. Esto se hace, por un lado, seleccionando los materiales a investigar y, por otro, dándole forma a la percepción de los investigadores. En nuestro ejemplo, los metadatos de la biblioteca influyen sobre el colectivo de científicos, porque los archivos METS/MODS ya contienen clasificaciones de tipos de texto. Estas etiquetas repercuten en la selección de textos o revistas hechas por los investigadores y en su comprensión de las características de los tipos de texto y, por lo tanto, estabilizan el canon literario.

En conclusión, para tomar conciencia crítica de este poder normativo y estar dispuesto a cuestionarlo o relativizarlo no basta con políticas de investigación que apelan al rendimiento y la prestación de servicios, aunque este sea un primer paso necesario. La Fundación Alemana de Investigación Científica (DFG) ya lo ha dado cambiando las reglas de la organización de la infraestructura bibliotecaria, pues,

además de la función de archivar libros impresos, ahora exige también servicios más flexibles que presten, sobre todo, ayuda y orientación a los investigadores en los nuevos entornos digitales. Sin embargo, para superar todos los retos, epistemológicos y praxeológicos que supone el cambio del estudio de libros al estudio de objetos digitales, habrá que mejorar aún la cooperación entre ambas partes: bibliotecas y archivos, de un lado, e investigación académica, de otro. Entender a la otra parte y las tareas prácticas con las que esta tiene que enfrentarse siempre es el paso decisivo para ello. Nuestro entorno virtual de investigación *Revistas Culturales 2.0* intenta dar buena cuenta de la necesidad de ese diálogo intercultural, interinstitucional e interdisciplinar.

REFERENCIAS BIBLIOGRÁFICAS

- BILBAO, F. (1866). *Obras completas*, vol. I. Buenos Aires: Imprenta de Buenos Aires.
- BOWKER, G. C. & LEIGH STAR, S. (1999). *Sorting Things Out. Classification and Its Consequences*. Cambridge, MA / Londres: The MIT Press.
- DASTON, L. (2008). "On Scientific Observation". *Isis* 99, 97-110.
- DRUCKER, J. (2011). "Humanities Approaches to Graphical Display". *Digital Humanities Quarterly* 5.1, s. p. Disponible en línea: <http://www.digitalhumanities.org/dhq/vol/5/1/000091/000091.html> [03/06/2019].
- EHRLICHER, H. (2016). "Revistas Culturales 2.0 o cómo investigar publicaciones periódicas en la era de las humanidades digitales". *Ínsula. Revista de Letras y Ciencias Humanas* 829-830, 32-34.
- FIGAROLA-CANEDA, D. (1922) (ed.). *Diccionario cubano de seudónimos*. La Habana: Imprenta Siglo XX.
- FLECK, L. (1980). *Entstehung und Entwicklung einer wissenschaftlichen Tatsache*. Frankfurt am Main: Suhrkamp.
- GRAY, J.; GERLITZ, C. & BOUNEGRU, L. (2018). "Data infrastructure literacy". *Big Data & Society* 5.2, 1-13.
- LEHMANN, J.; HUBER, E. & STODULKA, T. (2018). *H2020 Project K-PLEX: WP4 Report on Data, Knowledge Organisation and Epistemics*. Research Report: Freie Universität Berlin. Disponible en

- línea: <https://hal.archives-ouvertes.fr/hal-01761214> [03/06/2019].
- LYOTARD, J. F. (1988). *La condition postmoderne: rapport sur le savoir*. París: Les Editions de Minuit.
- MOLINA, C. A. (1984): *La revista "Alfar" y la prensa literaria de su época (1920-1930)*. La Coruña: Ediciones Nos.
- MORETTI, F. (2000). "Conjectures on World Literature". *New Left Review* 1, 54-68.
- ____ (2013). "Operationalizing or the Function of Measurement in Literary Theory". *New Left Review* 84, 103-119.
- ____ (2017). "Patterns and Interpretation". *Literary Lab Pamphlet* 15, s. p. En <https://litlab.stanford.edu/LiteraryLabPamphlet15.pdf> [03/06/2019].
- SCARONE, A. (ed.) (1942). *Diccionario de seudónimos del Uruguay*. Montevideo: Claudio García.
- TRILCKE, P. & FISCHER, F. (2018). "Literaturwissenschaft als Hackathon. Zur Praxeologie der Digital Literary Studies und ihren epistemischen Dingen". *Wie Digitalität die Geisteswissenschaften verändert: Neue Forschungsgegenstände und Methoden*, M. Huber y S. Krämer (eds.), s. p. Disponible en línea: http://www.zfdg.de/sb003_003 [30/06/2019].
- WESTBROOK, L. & SAPERSTEIN, A. (2015). "Rethinking the Measurement of Sex and Gender in Social Surveys". *Gender & Society* 29.4, 534-560.

Recibido el 13 de marzo de 2020.

Aceptado el 29 de abril de 2020.