

# La creación y gestión del archivo digital valleincliniano: *corpus manuscrito e impreso*

MARGARITA SANTOS ZAS (COORD.)

Grupo de Investigación Valle-Inclán, Universidade de Santiago de Compostela  
[margarita.santos@usc.es](mailto:margarita.santos@usc.es)

FRANCISCA MARTÍNEZ RODRÍGUEZ

Grupo de Investigación Valle-Inclán, Universidade de Santiago de Compostela  
[francisca.martinez@usc.es](mailto:francisca.martinez@usc.es)

ROSARIO MASCATO REY

Grupo de Investigación Valle-Inclán, Universidade de Santiago de Compostela  
[rosario.mascato@usc.es](mailto:rosario.mascato@usc.es)

## 1. EL FONDO DOCUMENTAL VALLEINCLANIANO DEL GIVIUS: LA PUESTA EN MARCHA DE UN ARCHIVO DIGITAL

Presentar al Grupo de Investigación Valle-Inclán de la USC (GIVIUS) es paso previo para explicar su trayectoria y objetivos, con los que se relaciona la construcción y gestión de su fondo documental y la puesta en marcha de su archivo digital.

Integramos el GIVIUS (<http://www.usc.es/es/catedras/valle-inclan/givius>) tres profesores de la USC, su núcleo estable desde hace casi dos décadas, cuatro doctores –tres de ellos colaboradores externos y el cuarto con un contrato posdoctoral– y cuatro doctorandas –dos de las cuales son contratadas con cargo a proyectos–. Es decir, somos once miembros, con formación teórica y experiencia compartida a lo largo de años, que hemos tratado de impulsar, desarrollar o culminar tareas difícilmente abarcables de modo individual.

Por otra parte, a partir del año 2002 me hice cargo de la dirección de la *Cátedra Valle-Inclán* (financiada en el marco del **“Programa Universidades”** del Grupo Santander), creada con el fin de apoyar y ampliar las actividades del GIVIUS. Así, las iniciativas de la *Cátedra*, que dispone desde 2009 de un portal en la Biblioteca Virtual Miguel de Cervantes (<<http://www.cervantesvirtual.com/portal/catedravalleinclinan>>), se han orientado, por una parte, a la formación de los integrantes del grupo de investigación; por otra, a la difusión de la investigación valleincliniana. En respuesta al primer objetivo nació el *Seminario Permanente*, impartido por especialistas en la obra del escritor gallego, proporciona a sus participantes la oportunidad de debatir temas y problemas relacionados con

los intereses investigadores del Grupo<sup>1</sup>. Para alcanzar el segundo, se puso en marcha uno de nuestros proyectos más acariciados, la *Biblioteca de la Cátedra Valle-Inclán*, Servicio de Publicacións e Intercambio Científico de la USC; una colección con sello propio destinada a la difusión de los estudios y ediciones valleinclinianos, que da cabida a obras propias y ajenas<sup>2</sup>. Por último, más de una década (2001-2013) llevamos impulsando el *Anuario Valle-Inclán/Anales de la Literatura Española Contemporánea* (Society of Spanish and Spanish-American Studies), que se ha convertido en un punto de referencia para los estudios valleinclinianos. Sumemos a todo ello, también como mecanismo de difusión, la organización de exposiciones (2008 y 2009) y congresos (6 hasta la fecha, el último en octubre de 2011), dedicados al escritor gallego.

En estos años hemos afrontado no pocos desafíos, obteniendo algunos logros particularmente destacables y desde hace algún tiempo hemos mirado hacia el horizonte que nos ofrecen las humanidades digitales.

Nada de la tarea llevada a cabo hubiese sido posible de no haber contado con financiación: cinco proyectos consecutivos del MEC, desde 1994, en que asumí la dirección del Grupo; y dos de la Xunta de Galicia, uno de ellos dirigido por Javier Serrano, han permitido consolidar nuestras líneas de trabajo. En la actualidad disponemos de un Proyecto subvencionado por el MEC (“La obra y el legado manuscrito de Valle-Inclán: ediciones y estudios críticos”, ref. FF12011-24130), incorporado a la Red-Investiga, y de una subvención de la Xunta de Galicia, concedida a GPC en el marco de las “Axudas do Programa de Consolidación e Estruturación de Unidades de investigación Competitivas” (ref. CN 2012/097).

Nuestro trabajo, cuyos resultados se han ido publicando en diversos medios especializados (ver la web del GIVIUS arriba citada), se asienta sobre un fondo documental y bibliográfico –prensa y obra editada–, en buena parte resultante de una indagación continuada en bibliotecas y hemerotecas tanto locales y nacionales como europeas y americanas; al igual que disponemos de la bibliografía primaria del escritor anterior a 1936; y contamos con una bibliografía secundaria actualizada, que permite abordar casi cualquier tema valleincliniano desde perspectivas diversas.

A este *corpus* impreso se han sumado en noviembre de 2009 los originales del desconocido Legado Manuscrito de Valle-Inclán, depositados en la USC por voluntad de sus propietarios<sup>3</sup>. Hasta esa fecha los autógrafos valleinclinianos se contaban con los dedos de una mano, aunque se han incrementado estos últimos años con algunas donaciones (diez cartas) y adquisiciones propias (poemas, cartas y tarjetas postales). A ese reducido patrimonio original se ha agregado

<sup>1</sup> Con un máximo de cuatro seminarios por curso académico, cuya duración no excede las ocho horas, en este foro de debate, que inauguró sus actividades en 2003, han intervenido dieciocho profesores de universidades o centros de investigación españoles, europeos y americanos, todos ellos reconocidos valleinclinianos.

<sup>2</sup> La colección tiene dos líneas editoriales: ediciones (facsimilares, anotadas, críticas...) y monografías y desde que inició su andadura, ha publicado ocho volúmenes, cuyos autores y títulos pueden verse en la web del GIVIUS, arriba citada. Baste decir que la colección se inauguró –y en esto ha sido pionera– con la edición facsímil del manuscrito de Valle-Inclán, *Mi Bisabuelo*, ed. de Jorge Devoto (2007); y el último de los títulos editados es un volumen colectivo: *Valle-Inclán y las artes*, eds. de Santos Zas, Serrano Alonso y De Juan Bolufer (2012).

<sup>3</sup> Para la historia de este excepcional Legado, su proceso de transmisión a la USC y la descripción de sus contenidos, ver Santos Zas (2008 y 2012).

el monumental archivo familiar Valle-Inclán/Alsina, que abre nuevas vías a la investigación valleincliniana, entre otras razones, porque significa contar con el primer eslabón de la cadena genética de algunas obras editadas, cuyas galeras corregidas también se conservan, así como –y es su principal novedad– los borradores de obras inéditas, de cuya existencia no se tenía noticia.

De las dos vertientes que conforman nuestro archivo, el actual *corpus* impreso incluye, en un arco de fechas que abarca desde 1888-1936, una serie de materiales que me limito a enumerar, ya que no puedo ni siquiera asomarme a la compleja casuística editorial ni al peculiar sistema de escritura y publicación del escritor (Santos Zas, 2013), con largas prehistorias, rastreables en la prensa y numerosas ediciones y emisiones de una misma obra, que presentan siempre variantes entre sí de mayor o menor calado y que, a los efectos que ahora interesa destacar, conlleva hacer acopio de múltiples versiones. En términos generales contamos con:

1. Más de trescientos textos, correspondientes a ediciones en vida del autor, antologías, prólogos y traducciones de sus obras, que en imágenes digitalizadas alcanzan a día de hoy la cifra de 36.346.
2. A este fondo editorial se agregan las publicaciones periodísticas, es decir, decenas de textos de creación aparecidos en la prensa, sean autónomos o dependientes de obras mayores.
3. De la constante indagación en la prensa proceden igualmente cientos de conferencias y discursos, entrevistas y encuestas y epistolario.
4. Contamos también con un fondo iconográfico, compuesto por fotografías, y caricaturas (que forman parte del archivo personal de Serrano Alonso). Y en el mismo terreno de la imagen, hemos reunido la mayoría de las ilustraciones y motivos ornamentales de las ediciones del autor (ver un amplio muestrario en el citado Portal de la Cátedra Valle-Inclán).
5. El GIVIUS en su indagación hemerográfica ha rastreado la recepción de la obra de Valle-Inclán en su tiempo, con atención particular a sus estrenos teatrales.
6. Por último, hemos reunido numerosas semblanzas y retratos literarios de Valle-Inclán, realizadas por periodistas, intelectuales y amigos.

En suma, estamos hablando de más de cuatro mil quinientos documentos distintos, que traducidos a imágenes digitales sobrepasan las 46.000. De modo que el total de los materiales impresos supera la cifra de 80.000 imágenes.

Desde muy pronto consideramos necesario que el material de prensa, conservado en archivos papel, se fuese incorporando a bases de datos para su posterior consulta y así se hizo ya en los años 90. Y en el caso de las ediciones valleinclinianas, antes incluso de esa fecha –siendo director del Proyecto Luis Iglesias–, utilizamos medios ópticos (Clarity) para la reproducción de imágenes, y OCR (“OmniPage”) para la conversión de dichas imágenes en texto. Este último trabajo se perdió en un intento posterior de trasvasarlo a nuevos sistemas y hubo que repetir el proceso íntegramente. Pues bien, si estos medios fueron hace dos décadas innovadores, es evidente que a estas alturas resultan insuficientes para dar respuesta a nuestras necesidades y objetivos.

Por lo que respecta al Archivo Valle-Inclán/Alsina, contiene lo que el propio escritor calificó de “pruebas de aguafuerte” (Valle-Inclán, 1993 [1924]: 160) ma-

nuscritos de trabajo, materiales en “ebullición” (Grésillon, 1994: 33). Y al decir manuscrito, lo utilizo en sentido genérico y uso el término borradores para señalar la fase redaccional de un autógrafo, un estado de su génesis, sea cual sea éste<sup>4</sup>. Pues bien, en este archivo familiar se conservan no sólo borradores de obras literarias de Valle-Inclán editadas, sino documentos rigurosamente inéditos: vocabularios o listas de palabras, cuadernos de viaje, notas documentales, esquemas, bocetos, resúmenes, ensayos redaccionales más o menos textualizados, versiones de una misma secuencia con tachaduras, sustituciones, adiciones y alteraciones o frecuentes variantes autoriales, copias autógrafas... de poemas, de piezas teatrales, cuentos, novelas, conferencias, pero también numerosas cuentas y facturas, que nos hablan de los detalles de la vida doméstica del escritor o de su relaciones con el mundo editorial... Se trata, en síntesis, del taller del escritor: más de cinco mil páginas autógrafas.

Pero este archivo guarda, además, galeradas corregidas por Don Ramón, ediciones revisadas y anotadas de su puño y letra y algunos mecanoscritos teatrales orientados a la representación escénica. A todo ello hay que añadir numerosos grabados y dibujos originales de artistas contemporáneos destinados al diseño gráfico de sus obras; más de un centenar de cartas originales, que incluyen en ocasiones a emisor y destinatario.

Es ingente la tarea que deparan estos fascinantes materiales<sup>5</sup>, que ayudan a resolver la inmensa laguna, que hasta su hallazgo existía respecto de los procesos de creación de la obra de Valle y sus estrategias de escritura –aproximarse a su *modus componendi*–, que habíamos basado en los pre-textos de muchas de sus obras, publicados en la prensa (con numerosas versiones), y en las ediciones sucesivas, con abundantes variantes textuales, publicadas antes de 1936, sobre las que se han apoyado las ediciones críticas hasta el presente.

<sup>4</sup> Sobre las definiciones dadas a los términos “manuscrito” y “borrador”, véase Grésillon (1994: 71 y 74). Por su parte, Javier Blasco (2011: 10), siguiendo a Irène Fenoglio (2007: 8-34), opta por el término *texto* (en cursiva) en oposición a ante-texto, para designar, en el primer caso, el “enunciado o conjunto coherente de enunciados orales o escritos tal como se recogen en la ‘copia final para la imprenta’ o en la edición impresa autorizada”; mientras que el ‘ante-texto’ remite a “las versiones que se contienen en todos los borradores previos a la copia para la imprenta”. Por otra parte utiliza ‘manuscrito’ en sentido genérico, al igual que se usa aquí, para definir el soporte sobre el cual se producen todas aquellas metamorfosis, resultantes del trabajo del escritor (psíquico, gestual y cognitivo), que deja huellas gráficas materializadas sobre el manuscrito. El ante-texto es el espacio propio de trabajo de la crítica genética.

<sup>5</sup> Afrontar su estudio no ha sido un proceso sencillo y en ocasiones nos hemos visto en la tesitura de tener que rectificar lo ya hecho, pues inicialmente (el primer convenio data de diciembre de 2007) dispusimos del fondo documental exclusivamente en soporte digital, cuyas condiciones no permitían una lectura y transcripción fiables. A finales de noviembre de 2009 se procedió al depósito de los originales del archivo Valle-Inclán/Alsina en la USC y el paso inmediatamente posterior fue acondicionarlos de acuerdo con las normas al uso para la conservación de este tipo de materiales, tarea meticulosa, que se prolongó hasta finales de 2010, de la que se ocuparon los responsables del fondo antiguo de nuestra Biblioteca Xeral, en cuya caja fuerte se custodian. Contar con financiación fue requisito imprescindible para proceder a su nueva y completa digitalización (marzo-junio 2012), esta vez acudiendo a medios profesionales, que garantizaban su fiel reproducción. Este proceso ha supuesto una ralentización del optimista plan de trabajo inicial, a la que también ha contribuido el paralelo proceso de aprendizaje de los integrantes del GIVIUS, orientado a afrontar la problemática teórico-práctica del estudio y edición de los manuscritos literarios, que en el caso de Valle-Inclán carecía de antecedentes, si exceptuamos la edición de *Mi Bisabuelo*, antes citada, depositada en la USC desde el 21 de junio de 2013 por voluntad de su propietario Jorge Devoto Valle-Inclán.

Si el estudio del archivo del escritor proporciona información muy valiosa, también suscita numerosos problemas: desde su datación, hasta los criterios para priorizar los varios borradores de una misma página, es decir, los diferentes estados redaccionales, cuyo orden de escritura se ignora. Cada borrador –soslayo ahora el concepto de *dossier* genético– exige un tratamiento individualizado que dé respuesta a los específicos problemas que conlleva. De ahí la necesidad de acudir a métodos, que desde focalizaciones atentas a estas realidades documentales y a sus concretos problemas editoriales, no sólo les confieran el valor que tienen para el conocimiento de los procesos de escritura de un autor, sino que proporcionen los recursos metodológicos idóneos para afrontar su carácter inestable, distintos aquellos de las herramientas que se utilizan para el estudio de la obra textualmente fijada.

El GIVIUS desarrolla líneas de investigación consolidadas, articuladas en torno al *corpus* impreso y manuscrito del escritor, que se engarzan en un mismo plan integrador, como diré después, al tiempo que afronta los retos de la aplicación en el campo de las Humanidades de recursos tecnológicos e informáticos *ad hoc*.

En este punto, y al igual que sucede con el *corpus* impreso, se ha realizado la digitalización del Legado Valle-Inclán, ya mencionada, con un resultado cuantitativo de 14.462 imágenes, que incluyen el recto y verso de cada hoja o cuartilla autógrafa de los diversos estados redaccionales que presentan los manuscritos de obras que finalmente dio a la estampa o que han permanecido inéditas; pero también la totalidad de las galeradas con correcciones autógrafas de obras que no llegó a editar; la *editio princeps* de otras con anotaciones manuscritas; el epistolario del escritor vilanovés; grabados y dibujos con destino al diseño de su *Opera Omnia*; y, por último, otros documentos: glosarios, notas de genealogía familiar, anotaciones históricas o esquemas de conferencias.

La digitalización de todo este material era absolutamente imprescindible para abordar, en primer lugar, su indexado y proceder después a su transcripción, estudio y edición. Recordemos que en el ámbito de la filología hispánica moderna y contemporánea, los editores de textos apenas encontraron formas para explorar y valorar mejor la gran riqueza material que contienen los archivos de autor. Los borradores y antetextos se suelen transcribir con las pautas de la crítica textual o ilustran simplemente algunas ediciones críticas, dando fe del paso del filólogo por los archivos del autor.

No es necesario recordar al respecto la labor realizada desde los años setenta por el ITEM (Institut des Textes et Manuscrits Modernes, París), que viene elaborando un protocolo de descripción, transcripción y estudio de los borradores modernos y contemporáneos, más acorde con la naturaleza de un texto “en proceso”. Es decir, los manuscritos de autor requieren dar respuesta a su específica problemática, diferente a la del texto impreso, acudiendo a la genética textual, que proporciona herramientas adecuadas para el estudio de los procesos de creación y sus particulares estrategias.

Nuestro desafío ante este inmenso archivo documental, su estudio y edición requiere no sólo disponer de conocimientos *ad hoc*, sino de medios para su uso y consulta con la mayor eficacia y rentabilidad. Y es aquí donde hemos acudido, sin renunciar a otros, a los recursos tecnológicos e informáticos en su aplicación al campo de las Humanidades, que presupone contar con la colaboración de técnicos en informática, pero también, en el caso de la catalogación, estudio y edición del fondo manuscrito original del escritor, requiere una formación específica de los integrantes del GIVIUS, que desde 2008 sigue siendo una meta,

que tiene su último reflejo en la participación de dos de sus integrantes en el taller “Humanidades digitales: enfoques y aplicaciones” (A Coruña, 8-10 de julio de 2013), pero cuenta con precedentes. El primero se presentó en 2009, recién depositados los manuscritos originales en la USC, con la invitación para participar en el “II Congreso Internacional de Manuscritos de escritores contemporáneos”, organizado por CiLengua. Un foro que supuso, por una parte, dar a conocer nuestro Archivo, y, por otra, debatir las propuestas más significativas en este campo con especialistas de varios países. A raíz de este Congreso surgieron otras propuestas del mayor interés para nosotros: en primer lugar, la colaboración con el seminario que impulsa el ITEM “Manuscritos hispánicos XIX-XXI”, y dirigen Bénédicte Vauthier y Fatiha Idmhand (enero, 2012); y meses después, la asistencia con dos integrantes del Grupo al “Workshop Digital Text/Genetic Document. Text Encoding Initiative and Genetique (TEI)”<sup>6</sup>, organizado por la prof<sup>a</sup>. Vauthier en la Universidad de Berna (septiembre, 2012), cuyo objetivo era la aplicación de los avances tecnológicos e informáticos al estudio y codificación de manuscritos, que se revela fundamental para difundir este patrimonio a gran escala, siendo en el ámbito hispánico una referencia la labor pionera del equipo de la Biblioteca Virtual Miguel de Cervantes, así como el proyecto “Biblioteca Digital Hispánica” de la Biblioteca Nacional de España.

La reunión de Berna fue un fecundo intercambio de experiencias en el terreno teórico, práctico y técnico sobre el manuscrito contemporáneo, que meses después tuvimos ocasión de verificar de la mano de Bénédicte Vauthier, profesora invitada en el máster de “Edición de textos” del Dpto de Literatura española, Teoría da Literatura e Lingüística Xeral, de la Facultade de Filoloxía (enero, 2013), al que asistieron miembros del GIVIUS. Este seminario se centró en el análisis y soluciones dadas a los problemas planteados por el complejo *dossier* genético de *Paisajes después de la batalla*, de Juan Goytisolo (Vauthier, 2012), cuyas pautas de trabajo constituyen un referente para el que nos ocupa<sup>7</sup>.

En nuestro proceso de aprendizaje, todavía abierto, han sido estos pasos esenciales para poder afrontar con responsabilidad y rigor la tarea encomendada por los Herederos de Don Ramón, conducente en último término a la edición de los materiales autógrafos, cometido que requiere su transcripción previa. Ello implica sistematizar una metodología de representación de estos materiales (las tres que se manejan habitualmente son la transcripción diplomática, semi-diplomática y diplomático-interpretativa), aunque este objetivo no excluye la realización de estudios concretos y su difusión, como venimos haciendo desde el momento mismo en que accedimos a este fondo, en revistas y foros internacionales, o su incorporación a las tesis de doctorado nuevas o en proceso. Por otra parte, no es necesario esperar a la transcripción íntegra de los materiales del citado Legado, pues algunos de los textos son unidades redaccionales acabadas y autónomas, que permiten su edición a corto plazo.

<sup>6</sup> Los investigadores (filólogos e informáticos) integrados en el programa “Text Encoding Initiative” (TEI), que concibieron el sistema de marcado de texto en XML-TEI –sistema que permite la producción de libros digitales–, tratan de adaptar estas herramientas informáticas al objeto de conseguir una transcripción fiel de los manuscritos contemporáneos.

<sup>7</sup> La prof<sup>a</sup>. Vauthier está trabajando en la construcción de un sitio web, en el que se puede seguir –ha tenido la gentileza de mostrárnoslo– el proceso de composición de *Paisajes después de la batalla* a través de un texto en movimiento. En la actualidad participamos en un proyecto conjunto, subvencionado por el FNS suizo.

Esta tarea editorial enlaza con una de las líneas de trabajo del GIVIUS más constantes: la elaboración de ediciones críticas y anotadas de la obra impresa del escritor. Inherente a esta labor es la recopilación de las sucesivas ediciones de cada uno de los textos del autor, al igual que sus numerosas versiones periodísticas, incorporadas, mediante la oportuna reelaboración, a obras mayores o a diversas colecciones, que, a su vez, modifican sus contenidos en sucesivas ediciones, hoy íntegramente digitalizadas.

De ahí también nuestro interés en continuar la indagación en la prensa en busca de todos aquellos textos de creación, aparecidos en periódicos y revistas, que forman parte de la génesis de las obras del escritor; continuar los estudios biográficos, que comporta ampliar la indagación hemerográfica en busca de materiales que los avalen. La abundante prensa periódica de la Edad de Plata y la foránea proporcionan información sobre la biografía de Valle-Inclán, su pensamiento estético y político: entrevistas, conferencias, retratos literarios, cartas y noticias varias, imprescindibles para alcanzar las labores propuestas.

Proseguiremos, de igual modo, el análisis de la recepción crítica contemporánea de la obra del escritor, la acogida de sus estrenos, fundamental para valorar sus éxitos y fracasos, el "impacto" en el panorama literario y cultural contemporáneo y su proyección innovadora.

Nuestro trabajo, que tiene asimismo una significativa vertiente hermenéutica, requiere la continua actualización de la bibliografía secundaria, constantemente incrementada, razón que justifica la preparación y puesta al día de repertorios bibliográficos, que se actualizan en cada uno de los sucesivos volúmenes del *Anuario Valle-Inclán*.

En suma: mientras que el Proyecto subvencionado por el MEC presta atención primordial a los manuscritos del Legado Valle-Inclán y a la coordinación de todas las líneas de investigación que se derivan del *corpus* manuscrito e impreso valleincliniano; la segunda de las líneas de trabajo del Grupo se viene desarrollando desde 2009 en el Campus de Lugo, a raíz de disponer de un Proyecto de investigación autonómico, subvencionado por la Xunta, dirigido por Serrano Alonso, que a partir de 2012 ha podido continuar su plan de trabajo en el marco del Plan de Consolidación de Grupos de investigación, de la Xunta de Galicia.

En nuestro ambicioso proyecto de trabajo ha sido objetivo prioritario y cumplido la completa digitalización de nuestro fondo documental y bibliográfico, como ya he dicho; ahora dos metas ocupan nuestro esfuerzo: continuar el indexado de todos los materiales valleinclinianos –manuscritos e impresos– recopilados a lo largo de los últimos veinte años; y dar cima a un Catalogador informático, que permitirá no sólo contener toda la información acerca de los documentos reunidos, sino ofrecer las citadas digitalizaciones y facilitar búsquedas integrales de todo tipo de información. Además, este Catalogador permitirá su utilización en red de uso restringido al GIVIUS, sin que descartemos en un futuro no lejano el acceso público a estos recursos.

Se ocupan de explicar a continuación este plan de trabajo, Francisca Martínez, doctoranda contratada del Proyecto, en "El archivo manuscrito digital valleincliniano"; y Rosario Mascato, titular de un contrato posdoctoral, en "Gestión y tratamiento del *corpus* impreso en el Archivo Digital Valle-Inclán". Ambas han estado en el origen y desarrollo de los respectivos archivos y darán cuenta de las herramientas informáticas de digitalización, catalogación, gestión y consulta de los materiales literarios, que responden al proceso de escritura y publica-

ción de la obra de Valle-Inclán, así como a su recepción crítica, sin soslayar la biografía del autor.

## 2. EL ARCHIVO MANUSCRITO DIGITAL VALLEINCLANIANO

Desde su creación, como se ha dicho, el GIVIUS ha ido haciendo acopio de lo que hoy es un voluminoso fondo bibliográfico y documental –impreso y manuscrito– que ya en la década de los 90 hizo necesaria la puesta en marcha de herramientas digitales, en concreto la primera base de datos en *access* para indexar entrevistas, cartas y conferencias valleinclinianas, creada, según las pautas de la dirección del GIVIUS, con la colaboración del profesor Gómez Segade, que ha sido útil hasta la puesta en marcha del archivo que ahora presentamos.

Por lo que se refiere al fondo manuscrito, hasta el año 2008 disponíamos de copias digitales de documentos autógrafos, raramente de creación (cartas, informes, memorias, dedicatorias, algunos poemas y un cuento..), que se vieron extraordinariamente incrementados, tanto en la vertiente creativa como en la documental, gracias al Legado Valle-Inclán Alsina (Fig. 1), cuyos originales están hoy depositado en la USC, según lo antes expuesto. A este archivo se han ido sumando otros materiales manuscritos adquiridos por el GIVIUS o donados a

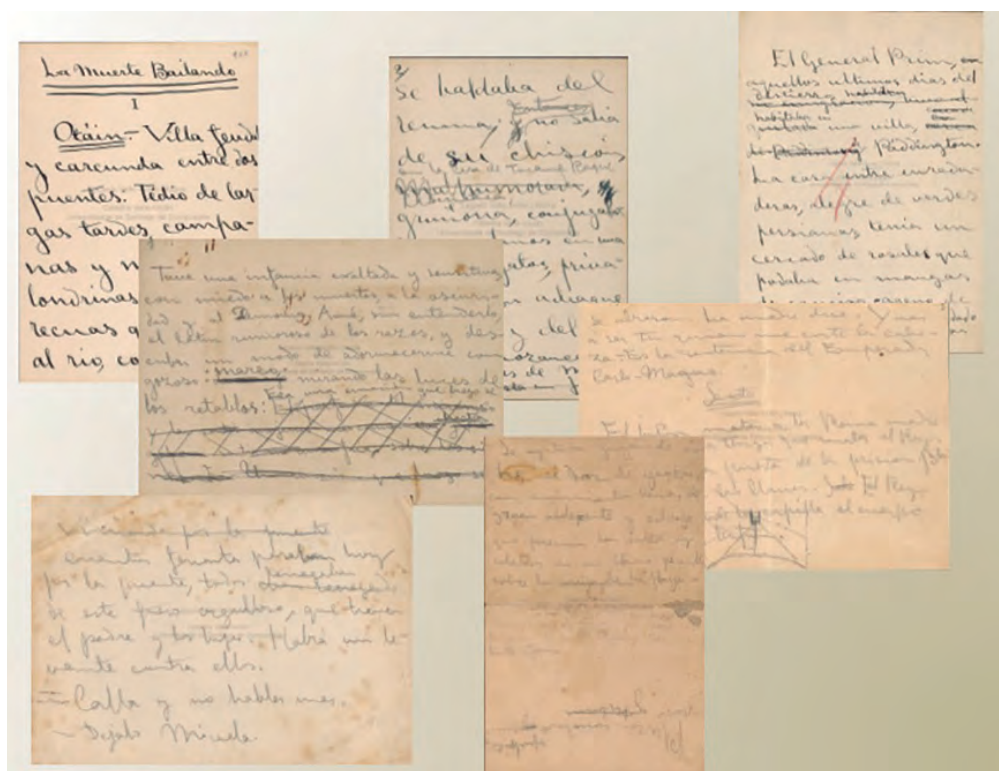


Fig. 1. Selección de páginas sueltas de varios borradores inéditos del Legado Valle-Inclán Alsina/Cátedra Valle-Inclán de la USC.



este por sus propietarios, de manera que para poder estudiarlos y editarlos se imponía la creación del Archivo Digital de Manuscritos Valleinclinianos, herramienta imprescindible para ofrecer el indexado de estos materiales en una base de datos relacional, que permitiese, en primer lugar, la ordenación, descripción y consulta de ese excepcional legado, para proceder, una vez culminada esta fase, a su integración en el Catalogador que se describe en el tercer apartado de este trabajo.

A tal fin, han sido determinantes en lo que se refiere a nuestro aprendizaje, los foros académicos arriba mencionados, en los que he participado con otros integrantes del GIVIUS, de los que destacaría por su carácter eminentemente práctico el *Workshop Digital Text/Genetic Document. Text Encoding Initiative and Genetic Edition*, celebrado en Berna (2012), bajo la coordinación de la profesora Bénédicte Vauthier. En esa ocasión las sesiones expositivas se combinaron con intensas sesiones prácticas sobre la aplicación de TEI, dirigidas, en nuestro grupo, por la profesora Elena Pierazzo, del Department of Digital Humanities del King's College de Londres, que generosamente nos facilitó numerosos materiales de trabajo y expuso la compleja praxis de codificación de manuscritos, ofreciéndonos una amplísima casuística de procesos de marcado TEI sobre textos de tipología genérica diversa. A esta experiencia se ha sumado ahora el workshop "Humanidades Digitales: enfoques y aplicaciones" (Universidade da Coruña, 8-10 de julio), orientado sobre todo al uso de herramientas digitales en el ámbito de las humanidades.

De cada uno de esos encuentros hemos extraído fórmulas y prácticas útiles para el tratamiento digital del fondo manuscrito de Valle-Inclán, a la vista de otras experiencias eficaces a la hora de manejar archivos similares en bases de datos que no sean meros repositorios, sino que sirvan de manera eficaz y rentable a la investigación del Grupo; es decir, que nos permitan indexar los materiales, formular campos con criterios genéticos y hacer uso de los metadatos para su consulta y posterior edición.

Para crear nuestra base, en este caso sin apoyo técnico externo, el primer paso era la indexación de materiales, y para ello hemos tomado como punto de partida el inventario de los documentos originales que conforman el Legado Valle-Inclán Alsina, elaborado por Joaquín del Valle-Inclán y revisado posteriormente con la profesora Santos Zas.

Este inventario consigna los materiales tal como fueron organizados por su propietario –Carlos del Valle-Inclán Blanco– y han llegado a nuestras manos, de modo que desconocemos su estado en origen. De hecho, el hijo del escritor realizó un primer intento de ordenar la totalidad de los borradores autógrafos con criterios temáticos (según se desprende de nuestra propia confrontación), ya que no están datados. De manera que los agrupó en carpetas comunes, numeradas y en su mayoría designadas con títulos –no siempre autoriales–, que acogen, a su vez, una serie variable de subcarpetas, que se enumeran en dependencia de la principal, cada una de las cuales mantiene cierta unidad de sentido, pero no con las que le preceden o le siguen. Quiero decir, que faltan engarces entre unas y otras, lo cual no siempre ni necesariamente significa que se hayan perdido, sino que en ocasiones se han ubicado en otras carpetas de contenido dispar, posiblemente como resultado de una lectura incorrecta o sencillamente porque, al tratarse de páginas sueltas en su mayoría, las posibilidades de que se altere el orden dentro y/o entre las subcarpetas, que integran cada carpeta, se multiplican.

Esta organización se ha respetado escrupulosamente al encargar a una empresa profesional la digitalización de la totalidad del Legado Valle-Inclán, proceso supervisado por la responsable del GIVIUS, que concluyó en junio de 2012. Esta nueva digitalización en verso y recto, que alcanza un alto grado de fiabilidad, constituye el segundo pilar para la construcción de la base de datos.

La creación de este archivo digital, que en último término deberá incorporarse al Catalogador “Archivo Digital del Grupo de Investigación Valle-Inclán”, para que alcance su pleno sentido y rendimiento, responde a los siguientes objetivos:

1. Facilitar los materiales indexados de acuerdo con la organización original, pero subsanando sus errores en la ordenación de contenidos.
2. Ofrecer una descripción material y textual de cada una de las unidades que conforman cada carpeta, creando campos *ad hoc* (Fig. 2), etapa ineludible en el proceso de estudio y edición de los manuscritos.
3. Proporcionar, en una segunda fase, la transcripción diplomática del documento, que deberá quedar respaldada por un archivo txt, necesario para la posterior incorporación al citado Catalogador.

**Legado Manuscrito Valle-Inclán**

Sinatura:

Denominación prooia:

Descripción:

Formato:

Soporte:

Instrumento:

Recto  Verso

Comienzo de página:

Final de Página:

Transcripción:

Texto en TEI:

Fecha de creación:

Revisión:






Fig. 2. Registro de FileMaker/Biblioteca de documentos, diseñado para el Archivo digital valleinclaniano.

- Complementariamente, crear una versión del texto, marcado con *Oxygen editor* (Fig. 3), en TEI (*Text Encoding Initiative*), consorcio que desarrolla y mantiene un estándar para la representación de textos en formato digital y ofrece un conjunto de directrices que especifican métodos de codificación de textos de lectura mecánica, sobre todo en las humanidades, las ciencias sociales y la lingüística. Desde 1994, las Normas TEI han sido utilizadas, además de por bibliotecas y museos, por editores e investigadores, pues aunque la adquisición de los conocimientos necesarios para su uso con pleno rendimiento no está exenta de complejidad, también creemos que las posibilidades que ofrece son enormemente rentables en función de compartir y trabajar en línea con una versión xml del documento (Fig. 4), que introduciría metadatos como parte de su información, viabilizando un sistema de búsquedas eficaz, y en consecuencia la elaboración de un *dossier* genético.

La codificación TEI permitirá aplicar estructuras genéricas para cada documento a través de la herramienta *ROMA* (Fig. 5), generador de esquemas determinados, con definiciones de elementos y atributos para cada tipología textual, en coherencia con la diversidad de materiales contenidos en el Legado Valle-Inclán Alsina.

En suma, tras el indexado, se obtiene un archivo por cada carpeta, que incluye tantos registros como imágenes estén contenidas en ella. Estos son fmp (FileMaker pro), si deseamos trabajar directamente en la base de datos, o pdf/Excel, sobre los que se podrán realizar búsquedas sin riesgo de modificar el contenido de los campos de la base.

Con el fin de agilizar este proceso, hemos optado por una base de datos con gestor integrado, creada a través de la aplicación FileMaker, una base de datos

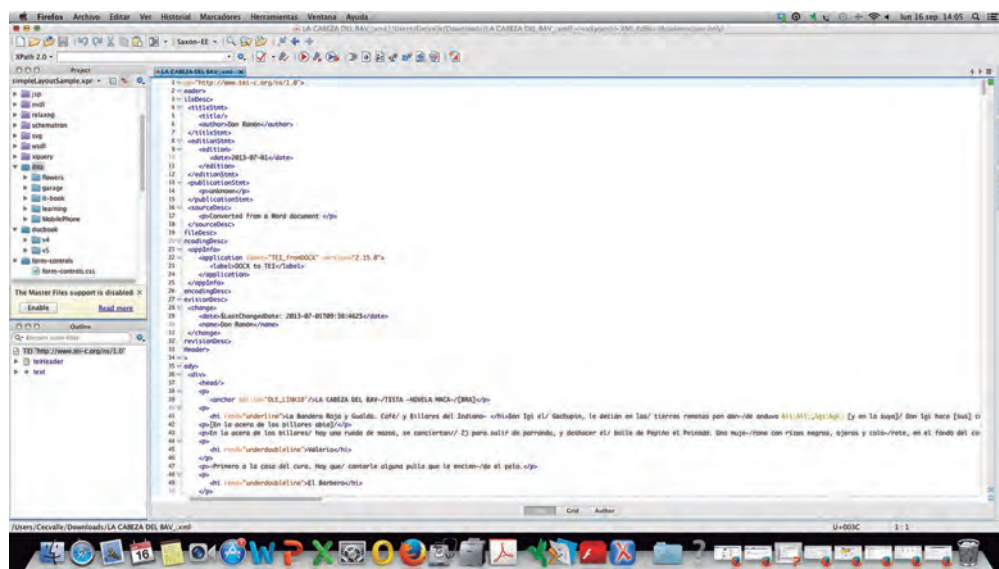


Fig. 3. Captura de pantalla de texto marcado con Text Encoding Iniciative (TEI), con el programa Oxygen Editor, modalidad de visualización "Text".

relacional que permite crear formularios, informes, etc., muy similar al modelo de Microsoft Access, en la que elegimos una solución prediseñada –Biblioteca de documentos–, modificando íntegramente los campos existentes para adecuarlos a las necesidades antes descritas.

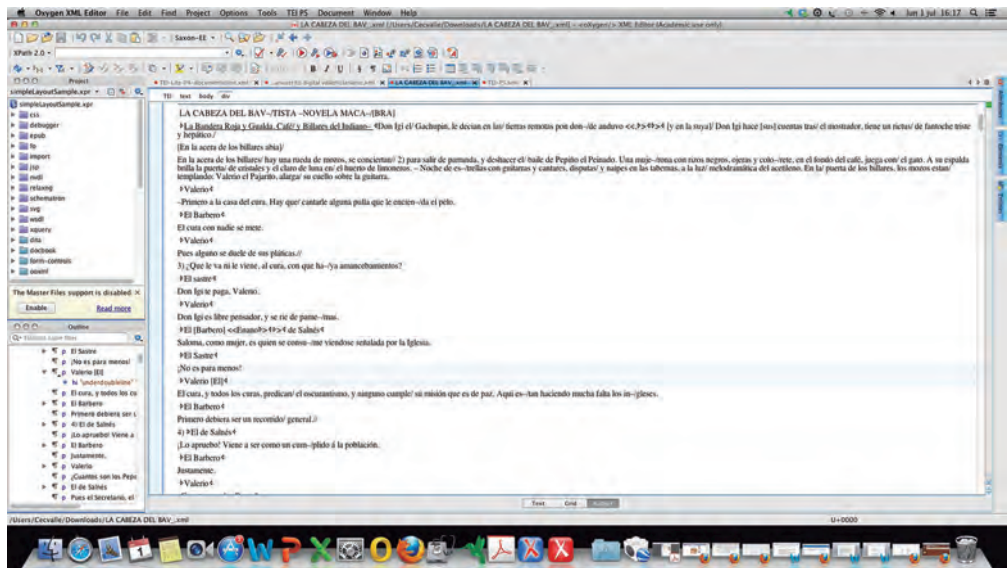


Fig. 4. Captura de pantalla de texto marcado con Text Encoding Initiative (TEI), con el programa Oxygen Editor, modalidad de visualización “Author”.

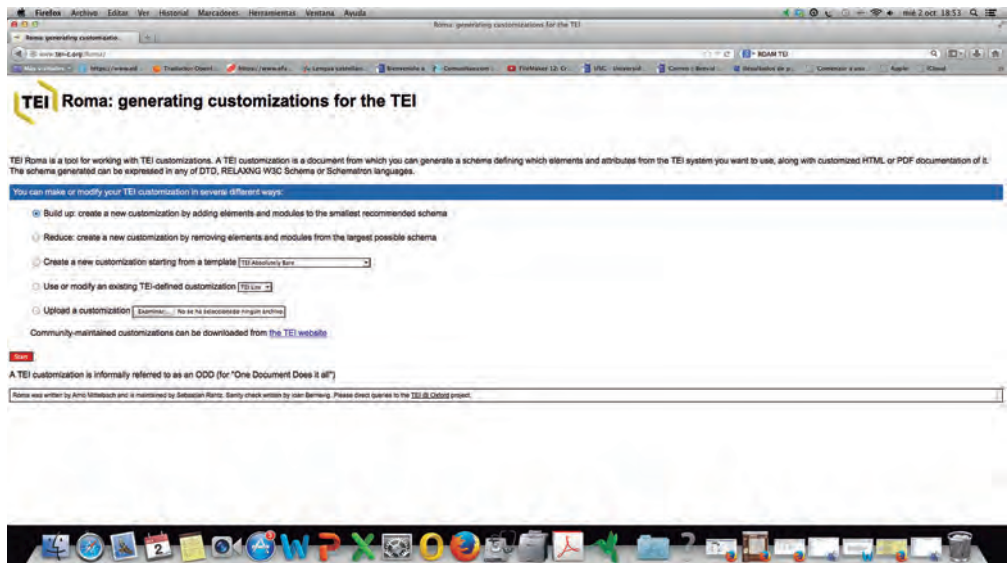


Fig. 5. Captura de pantalla de ROMA, página de inicio de generador de diseños para marcado TEI.

FileMaker puede importar datos de excel, archivos xml o fuentes ODBC y los informes generados pueden ser exportados en excel o PDF. En una fase más avanzada de nuestro trabajo utilizaremos Aperture, aplicación profesional de edición de imágenes digitales, que permite exportar los metadatos de las imágenes en un archivo de texto.

La versión servidor de la aplicación permitirá incorporar en futuro las bases de datos en el mencionado Catalogador. Hasta entonces, con la versión básica, tenemos la posibilidad de compartir las bases de datos con otros usuarios en una red local.

A la hora de elaborar los campos que figuran en cada uno de los registros, hemos acudido a criterios genéticos, que nos permitan una descripción material del documento. Así, contemplamos en nuestra base los siguientes campos:

1. **Signatura:** designación existente en el archivo original, que identifica cada carpeta e imagen en nuestro index.
2. **Denominación propia:** permite enmendar posibles denominaciones erróneas del punto anterior.
3. **Descripción:** en el registro inicial de cada archivo figura la adscripción genérica del documento: teatro o texto dialogado, narrativa, poesía, ensayo, glosarios, notas genealógicas, históricas, literarias...; se indica también el tipo de documento (borrador autógrafo, esquema, boceto, galerada...).
4. **Formato:** medidas del soporte establecidas con un medidor angular de editor de imagen.
5. **Soporte:** hoja suelta, cuartilla, cuaderno, bloc... Se consignan problemas que afectan al estado del soporte.
6. **Instrumento:** pluma o lápiz (negro, rojo, azul...).
7. **Espacio gráfico:** se indica su uso en recto y/o verso y recto.

En todo momento se puede visualizar la imagen que corresponde al elemento descrito, con posibilidad de exportarla fuera de la base.

En lo que al contenido del propio documento se refiere, sea cual sea su tipología, categoría y género, hemos diseñado cuatro campos:

1. **Palabras iniciales** de cada unidad digital, que ya figuraban en el inventario de los manuscritos originales.
2. **Palabras finales** de cada unidad digital, inexistentes en el citado inventario.

Ambos están destinados a la detección de incoherencias en la secuencialización del archivo.

1. **Transcripción**, según pautas descritas en el apartado 2 de los objetivos del archivo digital.
2. **Texto en TEI**, véase, asimismo, la explicación en el apartado 3 de los objetivos del archivo.

Estos dos últimos campos deben posibilitar al investigador una base textual fiable, en tanto uno y otro sistema están concebidos como punto de partida esencial para cualquiera de los estudios y modalidades editoriales que afrontemos en futuro.

El diseño por el que hemos optado, siempre perfectible, es sencillo pero eficaz, con la intención de agilizar su implementación en coherencia funcional con el modelo relacional del Catalogador y así tratar de lograr la creación de una macrobase de datos valleinclaniana, cuya aspiración es ofrecerse, en el contexto de las humanidades digitales y de la transferencia del conocimiento, como un modelo para la organización digital de los archivos de autor.

### 3. GESTIÓN Y TRATAMIENTO DEL *CORPUS* IMPRESO EN EL ARCHIVO DIGITAL VALLE-INCLÁN

Uno de los principales objetivos del Grupo de Investigación Valle-Inclán a lo largo de su historia ha sido ejercer una exhaustiva labor de recuperación de *documentación primaria* que, más allá, de la propia obra literaria del escritor, nos permitiese reconstruir su trayectoria en el marco del campo literario español de su época. Así, en los diversos estudios y trabajos publicados por los miembros del equipo, estos materiales se constituyen como eje para diversos abordajes metodológicos que, en unas ocasiones priman el análisis del proceso de elaboración literaria del autor (el denominado *taller del escritor* siempre considerado como *work in progress*); en otras, el estudio de las condiciones editoriales, institucionales e ideológicas que afectan a su producción en el marco del campo literario de inicios del siglo XX, con el afán de determinar el valor más o menos central o periférico de sus propuestas estéticas, identitarias o políticas con respecto a los repertorios en vigor en dicho campo. Independientemente de uno u otro enfoque, estos trabajos pretenden –en su conjunto– clarificar los límites existentes entre las verdaderas dimensiones de Valle-Inclán en la España de su tiempo, frente al conocimiento construido posteriormente por sus exégetas en el campo de la crítica valleinclaniana.

Para llevar a cabo esta tarea, es necesario trabajar con una extensa tipología de materiales que amplían la perspectiva generada por la obra literaria, en primera instancia, y ofrecen datos primarios en torno al desempeño de la labor del escritor: versiones de obras (previas o posteriores a la edición en libro); textos sobre estética y poética (bien en entrevistas, bien en artículos, conferencias o discursos); declaraciones públicas sobre asuntos políticos, culturales o sociales (como manifiestos o cartas públicas) a lo que se suman, por supuesto, primeras (y posteriores) ediciones de sus obras, al igual que antologías y colectáneas que recogen sus textos. Sin olvidar la propia recepción de época: reseñas críticas, artículos, materiales sobre sus estrenos teatrales e incluso parodias (por ejemplo, de sus poemas). A todo lo cual se suma un ingente material iconográfico que va desde centenares de caricaturas y una considerable colección de retratos y fotografías, a las ilustraciones que acompañan a sus obras (tanto en formato libro como en las páginas de la prensa). Y, claro está, otros materiales de primera mano, como epistolario, autógrafos o documentos personales.

En consecuencia, nuestro *corpus* de estudio se caracteriza por su variedad, extensión y riqueza en matices (Fig. 6). Se trata, en conjunto, de más de 4.500 documentos, de los que –una vez descontados los textos editados en libro, pequeños autógrafos y epistolario manuscrito– más del ochenta por ciento han sido recogidos de la prensa de la época (gallega, española, europea y americana), conformando así lo que hemos denominado como *corpus mediático* del escritor (Mascato, 2013).

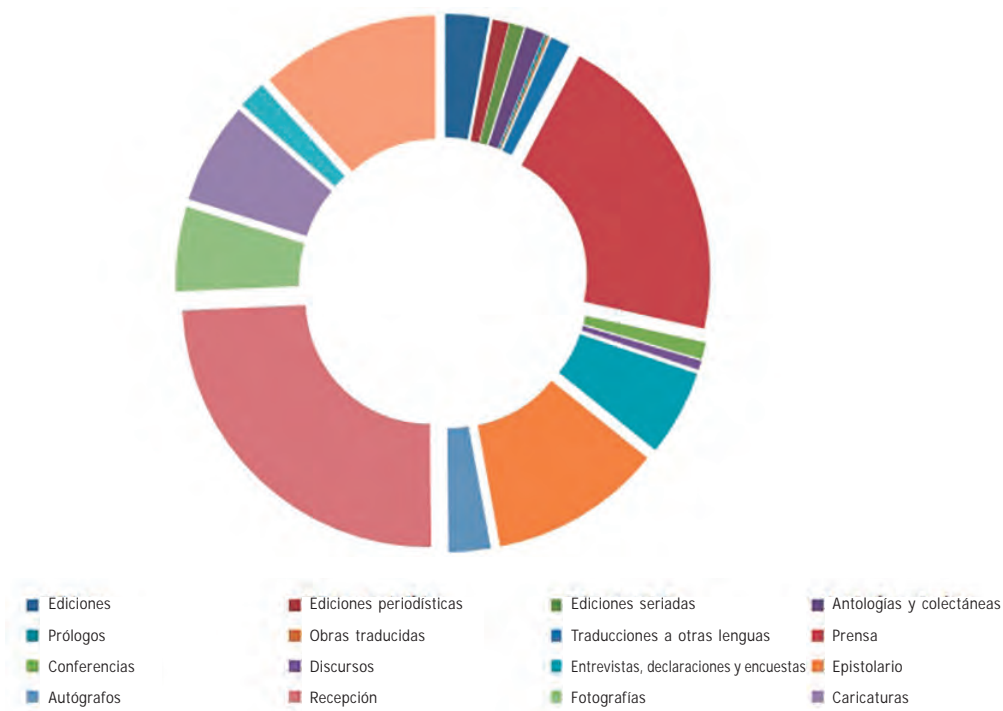


Fig. 6. Corpus impreso completo del Archivo Digital Valle-Inclán.

Este vasto *corpus* impreso requería, dada su complejidad y el estado precario en que se encuentra en muchos casos (libros con encuadernación deficiente, fotocopias con un tamaño o resolución inadecuados, microfilms de difícil acceso...), ser sometido a un proceso de ordenación y catalogación, para su posterior digitalización, una tarea que, con precedentes en el propio GIVIUS que se remontan a los años 90, alcanza su impulso definitivo en 2009 en el marco del proyecto “Valle-Inclán: a prensa e o sistema editorial” (2009-2012), dirigido por el profesor Javier Serrano Alonso, en la Facultad de Humanidades de Lugo, gracias a la financiación del Programa INCITE, de la Consellería de Industria de la Xunta de Galicia.

El proceso de digitalización de todos estos materiales, finalizado hace un año, consolidó las pautas de los procedimientos para el escaneado de textos (de acuerdo con los estándares actuales) –con un escáner adquirido *ad hoc* para el trabajo específico con libros (Plustek OpticBook 3600)–, la preparación de las imágenes individuales de cada página para obtener los mejores resultados de lectura –mediante *software* de edición de imagen (Adobe Photoshop CS3)–, la preparación de los PDF de cada obra o documento en su conjunto (vía Adobe Acrobat 9 Pro Extended) y el posterior tratamiento de cada documento con *software* de reconocimiento de texto (Abby Fine Reader 10 Corporate Edition), para la consecución de PDFs con función de búsqueda textual y señalización por marcadores, así como de documentos TXT pensados como base de gestión y búsqueda para diversas aplicaciones digitales.

En paralelo, a finales del año 2011 iniciamos el proceso de diseño conceptual y técnico de una macrobase de datos cuyos objetivos fundamentales eran:

1. Catalogar la documentación e información de ella derivada.
2. Indexar los documentos.
3. Disponibilizarlos para su consulta *online* y eventual descarga, de momento en un servidor con acceso restringido (al menos hasta 2016).
4. Habilitar la posibilidad de realizar búsquedas (por campo, categoría o incluso textuales).
5. Añadir nuevas tipologías de documentación.
6. Poder analizar esta información en clave textual, relacional, cuantitativa y gráfica en un futuro, mediante la utilización de nuevas aplicaciones y programas.

Ambos procesos confluyeron finalmente en la creación del **Archivo Digital del Grupo de Investigación Valle-Inclán**: una base de datos MySQL, exportable (por tanto) en lenguaje XML, y asimilable en la mayor parte de campos utilizados a los estándares internacionales (DublinCore, Marc21...) propios de bibliotecas y repositorios digitales de consulta pública<sup>8</sup> (Fig. 7).

Se trata de un macrocatalogador, en la actualidad dotado con diversas tipologías de usuario (superadministrador, administrador de unidad organizativa, catalogador, público), lo que permite simultanear el proceso de desarrollo y *upload* de documentos, con la gestión y tratamiento de los materiales, así como con su consulta.

En su estructura interna, el catalogador incluye diversas unidades organizativas, diseñadas en función de las características de las distintas secciones del *corpus*<sup>9</sup>: antologías y colectáneas, autógrafos, caricaturas, conferencias, dibujos, discursos, ediciones, ediciones periodísticas, ediciones seriadas, entrevistas, epistolario, fotografías, obras traducidas, prensa, prólogos y traducciones. Estas unidades funcionan como tablas relacionadas entre sí, que en algunos casos llevan asociadas otras con información que puede ser individualizada (y consultada) de manera transversal como parte de los intereses investigadores del equipo (lugares de publicación, periódicos y revistas, editoriales, imprentas, instituciones o agentes con que se relaciona Valle-Inclán, por ejemplo).

Cada registro incorpora los datos fundamentales del documento para su localización, así como una imagen del mismo, el archivo PDF correspondiente habilitado para búsquedas y el TXT sobre el que el propio catalogador genera la información textual. Todos estos datos están disponibles para el usuario final del Catalogador (Fig. 8).

Inicialmente ubicado en un servidor en pruebas, una de las incidencias de mayor relieve con que se encontró este proyecto fue el hecho de que la Universidad de Santiago carece, en estos momentos, de espacio en su servidor para

<sup>8</sup> En la actualidad, el proyecto se encuentra precisamente en fase de adaptación a estos estándares internacionales, llevando a cabo un proceso de revisión del catalogador.

<sup>9</sup> A estos efectos, hemos seguido, fundamentalmente, las pautas indicadas por Javier Serrano Alonso para su nueva versión de la *Bibliografía de Ramón del Valle-Inclán* –pendiente de edición–, que supone una ampliación y reformulación del *corpus* valleinclaniano en cuanto a la tipología y número de documentos en ella recogidos, frente a lo propuesto en Serrano Alonso y Juan Bolufer (1995).



The screenshot shows the public search interface of the Digital Archive of Valle-Inclán. At the top, there is a header with the USC logo (Universidade de Santiago de Compostela) and the text "ARCHIVO DIGITAL Grupo de Investigación Valle-Inclán" next to a portrait of Ramón del Valle-Inclán. Below the header, there are two main sections: a left sidebar and a main search area.

**Left Sidebar:**

- Menu Principal:** Inicio, Catalogador
- Menu de Usuario:** Logout

**Main Search Area:**

Archivo Digital Valle-Inclán

Unidad Organizativa: [ ] Lugar de Publicación: [ ] Año: [ ]

Filters (left):

- Antologías y Colectáneas
- Autógrafos
- Caricaturas
- Conferencias
- Dibujos
- Discursos
- Ediciones

Filters (middle):

- A Pobra do Caramiñal (A Coruña)
- Alicante
- Amsterdam
- Asunción (Paraguay)
- Bagnolet (Francia)
- Barcelona

Filters (right):

- 1866
- 1867
- 1868
- 1869
- 1870
- 1871
- 1872

Buttons: Marcar Todos / Desmarcar Todos (under each filter column)

Fecha: Desde: [ ] Hasta: [ ]

Búsqueda por campo: [ ] (separar palabras por comas)

Búsqueda en el PDF: [ ] (separar palabras por comas)

Serie: Nivel 1: [ Seleccione una opción ] Nivel 2: [ Seleccione una opción ]

Buscar

Fig. 7. Captura de pantalla del buscador público del Archivo Digital Valle-Inclán.

alojar una macrobase de datos de estas características y sus contenidos (que exceden en mucho los 25 Mb habitualmente disponibles a estos efectos). Esta situación, que no había sido contemplada en el *planning* inicial del proyecto, nos obligó a adquirir un servidor propio, al que la base fue trasladada –tras el período de construcción– en diciembre de 2012.

Desde esta fecha, el Archivo se encuentra en proceso de refinado técnico y de contenidos, en una exhaustiva revisión (en la que están implicados todos los miembros y colaboradores del Grupo), con el objeto de solventar cualquier incidencia que afecte a la efectividad y corrección en los procesos de catalogación y búsqueda: que no se produzcan fallos ni problemas de apertura de archivos, que el constante incremento del volumen de información no tenga incidencias sobre la velocidad y fiabilidad de las consultas, que la información sea indexada y jerarquizada de la manera adecuada con respecto a las prioridades del usuario en la consulta, etcétera.

Sin embargo, conviene indicar que un proyecto de esta entidad no está exento de problemas: en este caso precisamente derivados del trabajo con prensa histórica (grueso de nuestro *corpus*), cuya mayor dificultad reside en el proceso para la obtención del OCR (y, por tanto, de los archivos TXT, eje del trabajo de análisis y edición).

Si bien para el caso de aquellos materiales impresos en formato de libro, coleccionable o revista apenas hemos registrado dificultades de lectura, con esca-

Archivo Digital Valle-Inclán

Poemas de las Rosas. Rosa gitana [Rosa de Oriente]. Rosa de ultramar [Alegoría]



General	
Título	Poemas de las Rosas. Rosa gitana [Rosa de Oriente]. Rosa de ultramar [Alegoría]
Publicación	El Imparcial
Lugar (de publicación)	Madrid
Año	1918
Fecha	1918-06-24
Tomo, Nº, Páginas	p. 3
Género	Poesía
Serie	• Claves Líricas / El Pasajero
Ilustración	
PDF	<a href="#">Ver PDF</a>

Fig. 8. Resultado de una búsqueda en el Archivo Digital Valle-Inclán.

sas intervenciones en la corrección del texto, en el de la prensa diaria, concurren toda una serie de elementos que entorpecen la progresión del proyecto y nos han obligado a reconsiderar las prioridades y tareas inmediatas del equipo: el estado deficitario de la mayor parte de los periódicos y revistas originales, en lo referente a la calidad de impresión y el papel utilizado en la época; las dificultades de obtener una digitalización de determinadas cabeceras de prensa, en función muchas veces de su origen o de las instituciones que son depositarias de las mismas (es el caso de las bibliotecas cubanas, por ejemplo, donde la ma-

yoría de nuestros materiales han sido obtenidos en fotocopia); o no disponer de una digitalización adecuada, al haber sido conseguidas a través de repositorios digitales (el caso de la Hemeroteca Digital de la BNE) o de la solicitud de escaneados de uno u otro documento por parte de determinados archivos o hemerotecas, que cuentan con una política, medios técnicos y parámetros propios para este tipo de procesos.

Todo esto implica que, a pesar de utilizar las últimas versiones de software de reconocimiento de texto disponibles en el mercado (el mencionado Abby Fine Reader 10), los resultados estén lejos de ser satisfactorios, como se puede observar en los dos ejemplos aportados (Figs. 9 y 10).

Sin embargo, es esta una problemática común a otros proyectos de digitalización de prensa de los siglos XIX y XX. Incluso en aquellos casos en que los equipos de trabajo implicados cuentan con una capacidad de financiación mucho mayor a aquella de que puede disponer un grupo de investigación de las características del GIVIUS.

Un ejemplo reciente de ello lo ofrecía, en 2009, Rose Holley, directora del programa de digitalización de prensa histórica de la National Library of Australia:

In a sample of 45 pages to be representative of the libraries digitised newspaper collection 1803-1954, we found that raw OCR accuracy varied from 71% to 98.02%

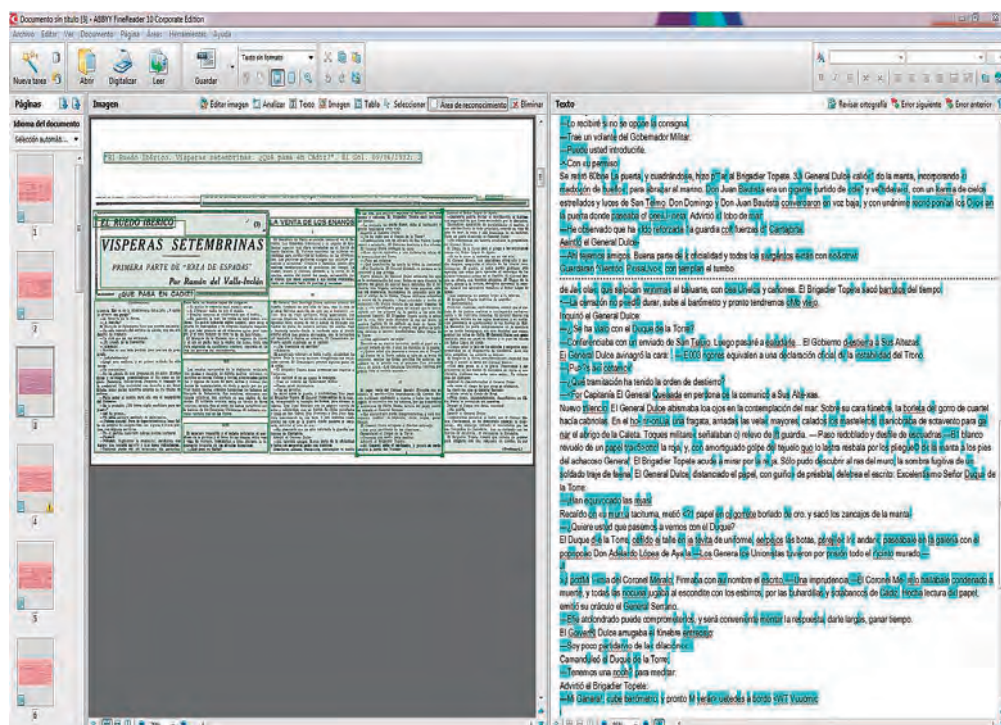


Fig. 9. Captura de pantalla de un proceso de reconocimiento de texto realizado con ABBY Fine Reader 10 – Corporate Edition, sobre un fragmento de “Visperas Septembrinas”, folletín perteneciente a la serie de El Ruedo Ibérico y publicado en El Sol, de Madrid (1932).

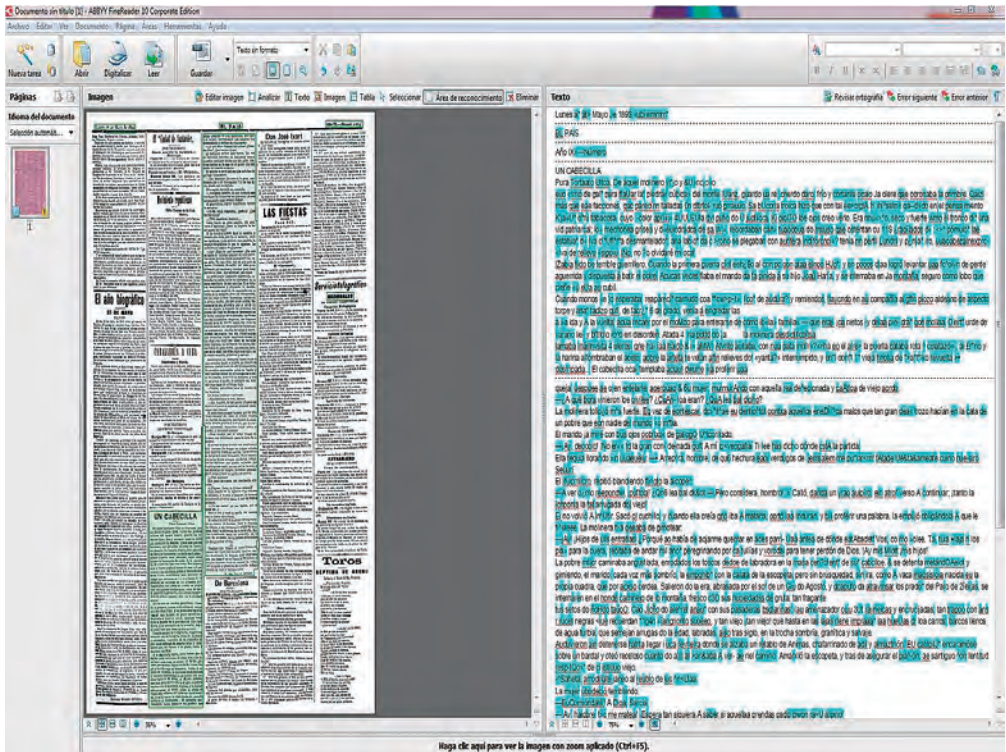


Fig. 10. Captura de pantalla de un proceso de reconocimiento de texto realizado con ABBY Fine Reader 10 – Corporate Edition, sobre el cuento “Un Cabecilla”, publicado en El País (1895).

(character confidence, no correction). At the 71% accuracy rate there would be 145 incorrect characters in an average paragraph of 500 characters. Looking at this the other way round means that 29% of the paragraph would be incorrect. This makes reading and accurate text retrieval difficult, even more so if the incorrect characters are not whole words but characters within many words. The question of what is acceptable has not been answered, but in speaking to other libraries and OCR contractors, it was generally agreed for historic newspapers that when we talk about good, average and bad OCR we mean:

Good OCR accuracy	= 98-99% accurate	(1-2% of OCR incorrect)
Average OCR accuracy	= 90-98% accurate	(2-10% of OCR incorrect)
Poor OCR accuracy	= below 90% accurate	(more than 10% of OCR incorrect)

Este es, por tanto, el gran problema al que se enfrenta un proyecto de nuestras características, para el que en la actualidad, las soluciones pasan no sólo por una mera cuestión de mejora de las condiciones técnicas de los materiales y el software utilizado para la obtención del OCR, sino –sobre todo– por un proceso de colaboración con el usuario del repositorio. Esto es, el “manual

method of humans correcting the mistakes of a machine", labor a la que nos hemos visto obligados a dedicar parte de los esfuerzos del GIVIUS a fin de conseguir el mayor grado deseable de corrección en los resultados obtenidos (muy especialmente en lo referente a las búsquedas textuales).

A pesar de estas dificultades, son varias las líneas de trabajo que se desprenden de la implementación de esta importante herramienta en el desarrollo de los objetivos del grupo de investigación:

Por una parte, a corto plazo, podremos examinar las condiciones de producción de la obra literaria en el campo literario, enfocando la trayectoria de Valle-Inclán en el mismo (su función y posición, de acuerdo con los parámetros establecidos por Bourdieu, 1992), a través de su relación con agentes e instituciones del mismo campo o de otros campos culturales y políticos. En este sentido, el análisis cuantitativo y relacional de los datos ya contenidos en nuestro catalogador nos permitirán, entre otras cosas, mapear la red de relaciones del escritor con su entorno, siguiendo el modelo de estudios anteriores referidos a casos literarios, como los de Samartin (2010) o Long y So (2013).

Por otra, tendremos la capacidad de profundizar, a medio plazo, en el estudio de la evolución estética y poética del literato a partir de datos empíricos y objetivos, obtenidos de secuencias textuales y concordancias a través de la aplicación del marcado semántico (TEI), lo que establecerá igualmente nuevos parámetros para la elaboración de ediciones críticas y genéticas hipertextuales. Para ello, resulta objetivo esencial del equipo obtener la transcripción (por medios técnicos y humanos) de la totalidad del *corpus* (manuscrito e impreso). El resultado, no menos ambicioso (por complejo) que su fase previa, nos permitirá visualizar en pantalla los diferentes estados creativos de una obra y la evolución, paso a paso, de Valle-Inclán, como productor de discursos e ideas.

En uno y otro caso, la tarea se antoja laboriosa, porque, entre otras cosas, pretendemos contribuir a un cambio de paradigma en el marco de la crítica (valleinclaniana), haciendo uso de esta nueva herramienta digital que es el Archivo Valle-Inclán para la mejor comprensión del hecho literario.

## BIBLIOGRAFÍA

- Blasco Pascual, Javier, *Poética de la escritura. El taller del poeta. Ensayo de crítica genética* (Juan Ramón Jiménez, Francisco Pino y Claudio Rodríguez), Valladolid/New York, Universidad de Valladolid/Junta de Castilla y León, Col. Cátedra Delibes, 2011.
- Bourdieu, Pierre, *Les Règles de l'art*, Paris, Éditions du Seuil, 1992.
- Burnard, Lou y Bauman, Syd, *TEI P5: Guidelines for Electronic Text Encoding and Interchange*, Charlottesville (Virginia), Text Encoding Initiative Consortium, 2013.
- Fenoglio, Irène, "Du texte avant le texte. Formes génétiques et marques énonciatives des pré-vision textuales", en *Langue française*, n° 155 (2007), pp. 8-34, <<http://www.cairn.info/revue-langue-francaise-2007-3-page-8.htm>>.
- Grésillon, Almut, *Éléments de critique génétique. Lire les manuscrits modernes*, Paris, PUF, 1994.
- Grésillon, Almut, *La mise en oeuvre. Itinéraires génétiques*, Paris, CNRS Editions, 2008.
- Hay, Louis et al., *Les manuscrits des écrivains*, Paris, CNRS, 1993.
- Koenig, Glenn, "The Origin of FileMaker Pro", <<http://www.dancing-data.com/filemakerhist.html>>.

- Holley, Rose, "How Good Can It Get? Analysing and Improving OCR Accuracy in Large Historic Newspaper Digitisation Programs", *D-Lib Magazine*, 15, 3/4 (March/April, 2009), <<http://www.dlib.org/dlib/march09/holley/03holley.html>>.
- Long, Hoyt y So, Richard, "Network Analysis and the Sociology of Modernism", *Boundary 2. An International Journal of Literature and Culture*, 40 (2) (2013), pp. 147-182.
- Marcos, M<sup>a</sup> del Carmen, "Los archivos en la era digital", <<http://goo.gl/7tw8Or>>.
- Mascato Rey, Rosario, *Valle-Inclán, poeta moderno no canonizado*, Universidade da Coruña, Servicio de Publicacións, 2013.
- Rodríguez, Javier, "FileMaker Pro 12: base de datos relacional", *Macworld España: la publicación para los usuarios del Mac, el iPhone y el iPod*, nº 224 (2012), pp. 94-95, <<http://www.idg.es/macworld/>>.
- Samartin, Roberto López-Iglesias, "Défices projetivos e estratégias de planificación cultural no campo editorial dum sistema periférico (Galiza: 1968-1978)", en M<sup>a</sup> Amparo Tavares Maleval y Laura Tato Fontaiña (eds.), *Estudos Galego-Brasileiros 4. Lingua, Literatura, Identidade*, A Coruña, Universidade da Coruña, 2010, pp. 255-276.
- Santos Zas, Margarita, "Los manuscritos de Valle-Inclán, existen", *Anuario Valle-Inclán VIII/ Anales de la Literatura Española Contemporánea*, 32/3 (2008), pp. 5-10.
- Santos Zas, Margarita, "Los manuscritos de Valle-Inclán: el taller del escritor", en Bénédicte Vauthier y Jimena Gamba Corradine (eds.), *Crítica genética y edición de manuscritos hispánicos contemporáneos*, Salamanca, Universidad, 2012, pp. 159-175.
- Santos Zas, Margarita, "Editar a Valle-Inclán: del manuscrito al impreso", en Ermitas Peñas Varela (ed.), *Perspectivas críticas para la edición de textos de literatura española*, Santiago, Servizo de Publicacións da Universidade, 2013, pp. 275-312.
- Serrano Alonso, Javier y de Juan Bolufer, Amparo, *Bibliografía general de Ramón del Valle-Inclán*, Santiago, Servizo de Publicacións da Universidade, 1995.
- Valle-Inclán, Ramón del, *Luces de bohemia. Esperpento* (1924), ed. Alonso Zamora Vicente, Madrid, Espasa-Calpe, Clásicos Castellanos, 1993.
- Vauthier, Bénédicte, *Juan Goytisolo, Paisajes después de la batalla. Preliminares y estudio de crítica genética*, Salamanca, Ediciones de la Universidad, 2012.
- Vauthier, Bénédicte y Gamba Corradine, Jimena (eds.), *Crítica genética y edición de manuscritos contemporáneos. Aportaciones a una poética de transición entre estados*, Salamanca, Ediciones Universidad de Salamanca, 2012.

## PÁGINAS DE INTERNET

- <<http://www.oxygenxml.com/>>.
- <<http://www.tei-c.org/index.xml>>.
- <<http://www.tei-c.org/Roma/>>.



## RESUMEN

A través de esta contribución pretendemos presentar el *Archivo Digital Valle-Inclán*, asociado a la trayectoria y líneas de trabajo del Grupo de Investigación Valle-Inclán de la USC (GIVIUS), desarrolladas a partir de un fondo documental –corpus manuscrito e impreso– cuya complejidad, riqueza y amplitud ha conducido a la construcción de dicho Archivo, cuyos objetivos, gestión y principales problemáticas se exponen de acuerdo con la tipología de los materiales que conforman dicho fondo documental y bibliográfico, reunido a lo largo de veinte años de investigación del GIVIUS.

*Palabras clave:* Archivo Digital Valle-Inclán, bases de datos, manuscritos, prensa histórica, herramientas digitales.

## ABSTRACT

Our aim is to present *Valle-Inclán's Digital Archive*, associated with the trajectory and work lines of the *Grupo de Investigación Valle-Inclán*, of the University of Santiago de Compostela (GIVIUS), and developed taking as a point of departure a wide, complex and varied documentary collection –here referred to as handwritten and printed *corpus*–. The objectives of this digital tool, its management and main problems found along the process of its creation, are exposed in the following pages, bearing in mind the kind of materials gathered in this documentary and bibliographical archive by the GIVIUS during the last twenty years of research activity.

*Keywords:* Valle-Inclán's Digital Archive, databases, manuscripts, historical press, digital tools.

