

La red de ontologías de POSTDATA (Poetry Standardization and Linked Open Data) para la descripción de Cancioneros

El proyecto POSTDATA (Poetry Standardization and Linked Open Data), financiado por ERC Starting Grant 2015 (ERC-2015-STG-679528, n.d.), se propone, con el ánimo de reducir la enorme brecha existente entre poesía y tecnología, acometer una importante transformación digital en el dominio de los recursos poéticos, de forma que puedan ser leídos e interpretados por máquinas para permitir su recuperación, indexación, vinculación y extracción de información adicional. Con sede en el LINHD (Laboratorio de Investigación en Humanidades Digitales) de la UNED, el proyecto cuenta con un importante equipo de investigadores para llevar a buen término este propósito.

POSTDATA es un proyecto multidisciplinar en lo que se refiere a su desarrollo, al combinar computación, ciencia de datos, filología, lingüística o historia literaria, lo que se refleja en el variado perfil de los miembros de su equipo y en sus resultados actuales; pero también mantiene esta amplitud y diversificación de intereses en lo que se refiere a su audiencia. Por una parte, el filólogo se acerca a las herramientas desarrolladas en el marco del proyecto para estructurar y analizar de forma precisa el material poético, para llevar a cabo una edición crítica o para recuperar datos y extraer información valiosa para su investigación. Los usuarios no especializados, ajenos al mundo académico, pero con aficiones literarias, pueden recurrir a POSTDATA para elaborar propuestas formativas, musicales u orientadas al turismo, o, por citar una de las últimas propuestas, la iniciativa #confinedverses, que tiene como objetivo recopilar y analizar poemas y letras de canciones en torno al COVID-19. Por último, el programador o el especialista en cualquier campo de la computación puede contribuir al desarrollo de herramientas de análisis, partiendo de lo ya realizado, en código abierto, y disponible a través de Github.

PoetryLab¹ y Ontologies² son las dos ramas del proyecto o, si se prefiere, las dos líneas de desarrollo, perfectamente integradas. La primera tiene como objetivo el desarrollo de herramientas de análisis básico de objetos poéticos, pero de gran complejidad computacional. Así, cuenta con una aplicación cuya finalidad es la escansión de un poema, denominada Rantamplan, desarrollada en Python, o Jollyjumper, desarrollada en el mismo lenguaje, que identifica los casos de encabalgamiento en un poema. Averell, por citar otra herramienta, permite la anotación de corpus poéticos y la obtención de un archivo JSON con etiquetas homólogas a las clases y propiedades de la red de ontologías POSTDATA. Ontologies, por su parte, aproxima la poesía a las tecnologías de la web semántica, que dota a sus recursos de anotaciones semánticas legibles por máquina, lo que permite establecer relaciones de significado entre ellos y realizar operaciones de inferencia para obtener informaciones adicionales. Las ontologías, con sus lenguajes RDF o OWL³, y Linked Open Data (LOD)⁴ son algunos de los estándares que están en la base del proyecto y que configuran el mundo de la web semántica, ese sueño preconizado por Berners-Lee (2001), pero todavía en situación emergente en muchos campos.

Desde la perspectiva de una biblioteca histórica interesa especialmente esta última vertiente de POSTDATA, como punto de partida para el desarrollo de herramientas útiles para una aproximación más en profundidad, en particular, a sus materiales poéticos. En efecto, la Real Biblioteca, al igual que otras colecciones patrimoniales, dispone de una gran cantidad de poemas recogidos en cartapacios o cancioneros, poemarios unitarios, poemas sueltos o insertos en obras de diversa naturaleza, que necesitan una herramienta especializada de análisis y clasificación, que identifique cada una de las unidades poéticas. Tomando como ejemplo la



¹ PoetryLab: <http://postdata.linhd.es/results/poetrylab>.

² Network of Ontologies: <http://postdata.linhd.uned.es/results/ontologies/network-of-ontologies/>.

³ RDF: <https://www.w3.org/RDF/>; OWL: <https://www.w3.org/OWL/>

⁴ LOD: <https://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData>.

AVISOS

RB, es bien conocida su colección de cartapacios desde los estudios de Menéndez Pidal (1914), o el Cancionero de Palacio, que cuenta con importantes estudios y ediciones, desde la de Barbieri de 1890.

Las herramientas de descripción de que se dispone, MARC⁵ o FRBR⁶ permiten aislar, de acuerdo con sus objetivos, los elementos identificativos de la naturaleza puramente bibliográfica del recurso, dejando al margen de su naturaleza literaria. POSTDATA parte del potencial descriptivo bibliográfico de esos recursos, y amplía el dominio a las entidades de tipo poético transmitidas por estos materiales bibliográficos, es decir, trasciende la naturaleza bibliográfica de los recursos poéticos para abordar también el análisis de su naturaleza textual.

En efecto, la red de ontologías de POSTDATA, tal como está configurada en la versión actual, cuenta con cuatro vocabularios de dominio principales: **posdatata-core**, **postdata-transmission**, **postdata-prosodicElements** y **postdata-structuralElements**. Los dos primeros, **core** y **transmission**, se ocupan de la descripción de la obra y del objeto bibliográfico que la transmite y son herederas estructuralmente de Functional Requirements of Bibliographic Records (FRBR) y, dentro de esta familia se ha optado por la reformulación FRBRoo, extensión oficial de CIDOC CRM, en la versión disponible de POSTDATA. Sin embargo, la entrada en escena de Library Reference Model (LRM)⁷, nuevo miembro de la familia FRBR, que se propone como el desarrollo consolidado de las modalidades anteriores y que cuenta ya con su codificación RDF, plantea dudas en cuanto a la selección de FRBRoo y aconseja la adopción de la nueva versión⁸.

En cualquier caso, es común a todos los modelos el mantenimiento de la jerarquía FRBR original, que, de mayor a menor nivel de abstracción, conduce de la obra (**Work**) al ejemplar (**Item**). La ontología **postdata-core** se centra en los dos primeros niveles, que redefine como **PoeticWork**, subclase de **frbr:Work**, y **frbr:Redaction**, subclase de **frbr:Expression**. **PoeticWork** es la creación artística en su concepción más abstracta, que se realiza en un texto concreto, conceptualizado en la clase **Redaction**. La **Redaction**, en efecto, es la clase que conceptualiza la naturaleza textual del poema.

Por otra parte, la segunda ontología, denominada **Transmission**, integrada en la red de ontologías POSTDATA, se ocupa de la manifestación física, es decir, de la materialización (*embodiment*) de la **Redaction**. A la **Transmission** corresponden dos niveles de abstracción: el denominado **Source**, que equivale a la **Manifestation** de las FRBR, y el denominado **Exemplar**, que se corresponde y es subclase de **frbr:Item**. La **Source** es el artefacto que transmite el texto, y puede tratarse de la copia ideal de una serie -si se refiere a una edición impresa- o, en el campo de la transmisión manuscrita, al artefacto único que transmite el texto. El **Exemplar**, por su parte, es la entidad física tangible. Evidentemente, en el ámbito de la transmisión manuscrita la diferenciación entre ambos niveles es más complicada, dado que en este caso se trata de una manifestación de carácter individual, si bien la clase que modela a cada uno de ellos dispone de sus propiedades específicas. Mientras que **Source** recoge los atributos necesarios para identificar las menciones de título o responsabilidad, la extensión, dimensiones o tipo de soporte, el **Exemplar** se ocupa de las propiedades que localizan físicamente el artefacto junto con la historia de esa localización: ubicación, signatura, datos de adquisición o información sobre su procedencia o custodia.

Estos niveles de abstracción están definidos partiendo del principio de encapsulación, de modo que la relación entre ellos se establece a través de una interface bien definida, que se concreta en una determinada propiedad de clase. Por otra parte, se establece una estructura de capas, de modo que una determinada capa solo conoce (se relaciona con) la inmediatamente inferior. La relación entre el primer nivel, **PoeticWork**, y el segundo, **Redaction**, se establece por medio de la propiedad de objeto **isRealizedThrough**; **Redaction** se relaciona con **Source** a través de **Source**, subpropiedad de **isEmbodimentOf**; y, finalmente, la **Source** se relaciona con el **Exemplar** a través de la propiedad **Exemplar** o, inversamente, **isExemplarOf** (Figura 1: De **PoeticWork** a **Exemplar**).

⁵ MARC: <http://www.loc.gov/marc/marc.html>.

⁶ FRBR: <https://www.ifla.org/publications/functional-requirements-for-bibliographic-records>.

⁷ IFLA LRM: https://www.ifla.org/files/assets/cataloguing/frbr-lrm/ifla-lrm-august-2017_rev201712-es.pdf.

⁸ Para una evaluación comparativa de estas variantes de FRBR y de BIBFRAME puede verse (Zapounidou, Sfakakis, & Papatheodorou, 2017)

AVISOS

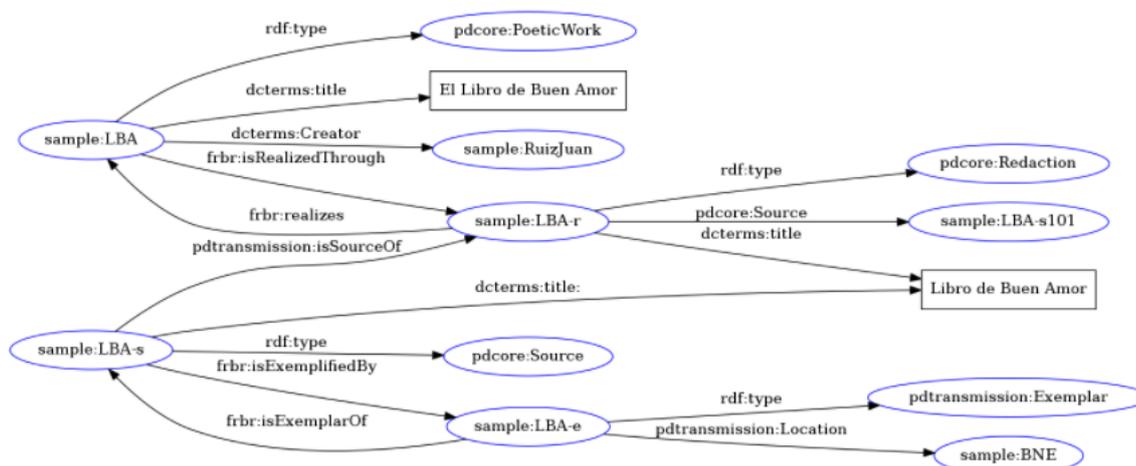


Figura 1: De PoeticWork a Exemplar

Para la descripción pormenorizada de cada uno de los atributos de estas clases se podrá hacer uso de RDA⁹, por su precisión semántica y su fina granularidad. En efecto, el modelo FRBR y, en concreto su modalidad LRM, cuenta con mecanismos perfectamente diseñados para su expansión. Por ejemplo, el atributo *Note* de LRM podría servir de atributo padre para todos los tipos específicos de notas definidos en RDA (Riva, 2017, p. 8). En la Figura 2, a modo de ejemplo, se ha descrito de forma abreviada un cancionero impreso de Juan del Encina (RB I/B/15) en el que se han utilizado las propiedades *rdam:P30156* 'has title proper', *rdam:P30134* 'has title of manifestation', *rdam:P30358* 'has printer person', *rdam:P30088* 'has place of publication', *rdam:P30011* 'has date of publication', *rdam:P30197* 'has bibliographic format', *rdam:P30254* 'is manifestation described in', dentro de la clase *pdcore:Source*, subclase de *frbr:Manifestation*. Sin embargo, para menores niveles de detalle se acepta el uso de las propiedades de *dcmi-terms*¹⁰.

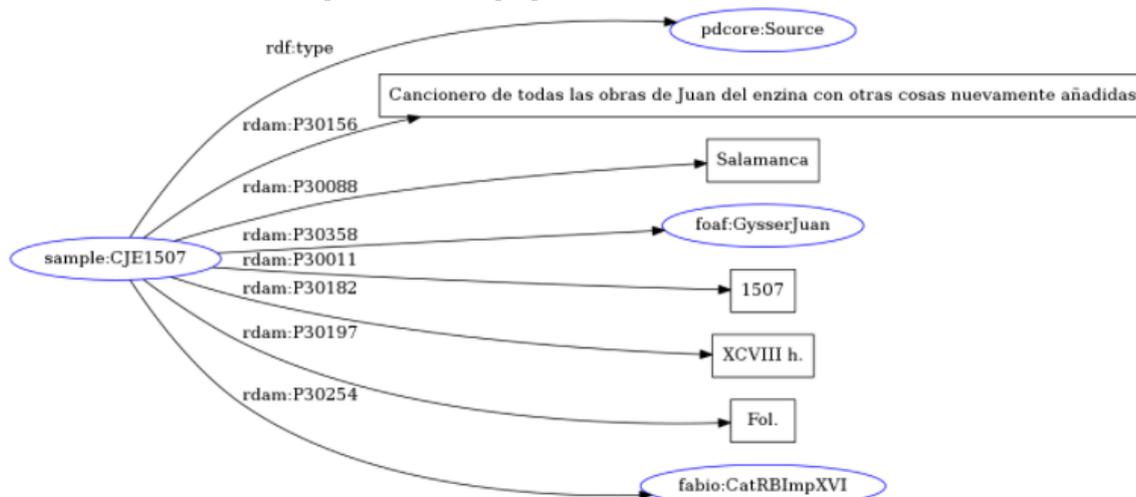


Figura 2: Descripción parcial de un impreso con RDA

Para personas y entidades corporativas, por una parte, y para conceptos, objetos, eventos y lugares, por otra, se recurre, respectivamente, a las clases de los grupos 2 y 3 de la FRRB, que integran, respectivamente, los vocabularios FRAD¹¹ y FRASD¹². Sin embargo, se aconseja el tratamiento unificado que se realiza en LRM, lo que constituye una de sus grandes aportaciones.

Por lo que se refiere a **Redaction**, entidad textual y clase nuclear en POSTDATA, se define a través de dos propiedades de tipo objeto obligatorias, **PoeticWork**, que identifica la obra, en este caso la composición poética, y **Source**, que identifica la fuente de la que se toma. En la Figura 3 se observa cómo **sample:LBA-r** es **Redaction** de la **PoeticWork**, **sample:LBA**, y que de entre las dos fuentes (**Source**) de que dispone, **sample:LBA-s1** y **sample:LBA-s2**, toma su texto de la primera.

⁹ RDA elements sets: <http://www.rdaregistry.info/Elements/>.

¹⁰ DCMI Metadata Terms: <https://www.dublincore.org/specifications/dublin-core/dcmi-terms/>.

¹¹ FRAD Functional Requirements for Authority Data: <https://www.ifla.org/publications/functional-requirements-for-authority-data>.

¹² FRASD Functional Requirements for Subject Authority Data: <https://www.ifla.org/node/5849>

AVISOS

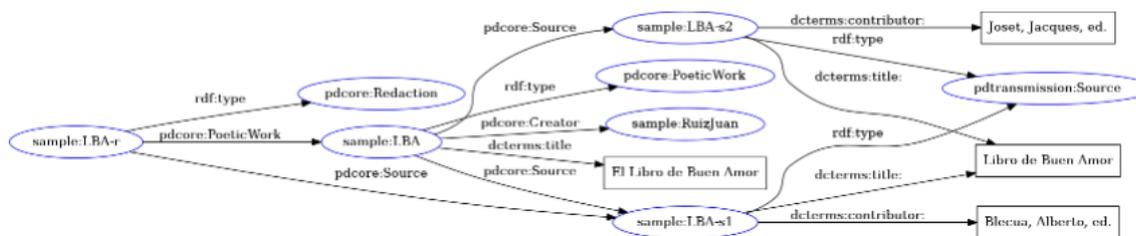


Figura 3: Redaction

Las relaciones entre parte y todo constituyen uno de los aspectos más discutidos en los entornos de trabajo y desarrollo de la FRBR (O’Neill, Cato, Goosens, Tillet, & Kuhagen, 2011; O’Neill & Zumer, 2015) y son especialmente pertinentes en los materiales poéticos, ya que, entre ellos, podemos identificar poemarios concebidos unitariamente o poemas sueltos formando parte de recopilaciones, debidas a la voluntad de un editor o de un poseedor. Este segundo tipo es tal vez el más frecuente en el caso que nos ocupa. Blecua (1983, p. 190), gran conocedor de la transmisión manuscrita e impresa así lo declara:

En general [las ediciones sin intervención del autor] son muy frecuentes en los siglos XVI y XVII, y en particular en la lírica y el teatro, yua que los poetas no fueron muy aficionados a publicar unas obras que una pujante tradición manuscrita y oral podía difundir suficientemente.

El tratamiento en ambos supuestos debe ser diferente. En el primer caso el poemario como conjunto es objeto de una categorización como **PoeticWork**, ya que así ha sido concebido originalmente por su autor. En el segundo caso, que podría corresponder a un cancionero, la concepción del conjunto está en el nivel de la transmisión, en concreto, de la **Manifestation** o, en nuestro caso, de su subclase **Source**. En este supuesto, cada una de las composiciones constituye un **PoeticWork**, y se reserva al nivel de la transmisión el tratamiento del volumen compilatorio y el establecimiento de las relaciones parte – todo.

Si se toma como ejemplo el *Libro de Buen Amor* y se considera un poemario de concepción unitaria, para las partes componentes se utilizarán las clases **PoeticWorkComponentPart**, en el nivel **Work**, y **SourceComponentPart**, en el nivel **Manifestation**, que, definidas como subclases de las correspondientes **PoeticWork** y **Source**, representan, respectivamente, una composición poética que es parte de una obra unitaria, y la descripción de la parte exacta que ocupa en la fuente. En la Figura 4 se observa cómo la composición poética *sample:LBA-p1*, que tiene por título “Enxiemplo del león e del mur” es parte de **PoeticWork** *sample:LBA*, que tiene como fuente la **SourceComponentPart** *sample:LBA-s101*, que corresponde a los vv. 1425-1439, localizados en las págs. 362-365 de la fuente *LBA-s1*.

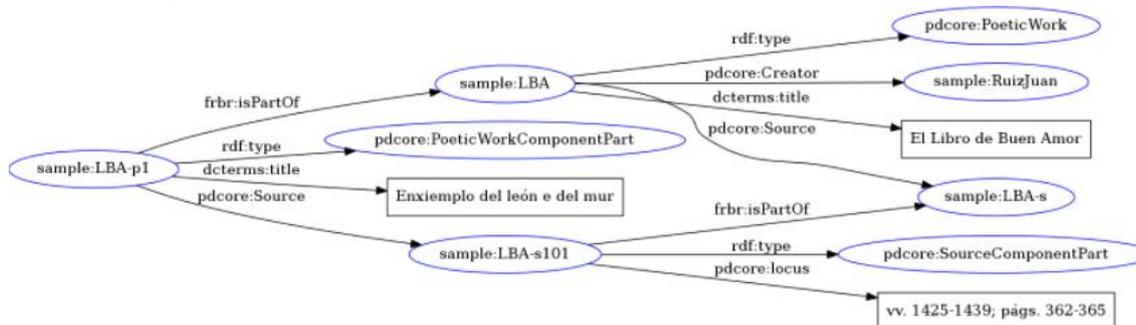


Figura 4: PoeticWorkComponentPart y SourceComponentPart

Para el segundo caso, composiciones poéticas en agregaciones realizadas en el nivel de la Manifestation, obra por tanto de editores o compiladores, se ofrece el ejemplo reflejado en la Figura 5. El poema, en este caso de Hurtado de Mendoza, es considerado como **PoeticWork** que tiene como **Source** una parte componente, **pdtransmission:SourceComponentPart** de un Cancionero, **pdtransmission:Source**, es decir, la agrupación de poemas se define en el nivel de abstracción, **Manifestation**, en el que es concebida.

AVISOS

- O'Neill, E., Cato, A., Goosens, P., Tillet, B., & Kuhagen, J. (2011). *Final Report of the Working Group on Aggregates*. Oslo.
- O'Neill, E., & Zumer, M. (2015). Aggregates as Manifestations: O'Neill & Zumer Proposal. Retrieved January 12, 2020, from <http://www.ifla.org/files/assets/cataloguing/frbrg/aggregates-as-manifestations.pdf>
- Riva, P. (2017). *Bulding RDA using the FRBR Library Reference Model*. 1–12.
- Zapounidou, S., Sfakakis, M., & Papatheodorou, C. (2017). Representing and integrating bibliographic information into the Semantic Web: A comparison of four conceptual models. *Journal of Information Science*, 4(43), 525–553. <https://doi.org/10.1177/0165551516650410>

AVISOS