

## VITIS CATALANA: PORTAL ESPECIALITZAT SOBRE LES VARIETATS DE VINYA CATALANES

Jordina Escala, bibliotecària, Centre de Documentació VINSEUM, c. Palma 24, baixos, 08720, Vilafranca del Penedès, tel. 938 923 204, [jescala@vinseum.cat](mailto:jescala@vinseum.cat), [www.vinseum.cat](http://www.vinseum.cat).

Marc Folia, consultor, Nubilum SL, Gran Via de les Corts Catalanes, 575, 1r 1a, 08011, Barcelona, tel. 93 100 30 94, [marc.folia@nubilum.es](mailto:marc.folia@nubilum.es), [www.nubilum.cat](http://www.nubilum.cat).

### Resum:

*Vitis catalana* ([www.vitiscatalana.cat](http://www.vitiscatalana.cat)) és un portal especialitzat en les varietats de vinya als territoris de parla i cultura catalana. Ofereix informació històrica i terminològica sobre aquestes varietats, basant-se en el buidatge de documents en forma de cites de presència. El projecte s'ha desenvolupat amb una aplicació basada en el model de dades RDF i prenent com a referent estàndards com el CIDOC CRM i l'SKOS. L'ús de *GeoNames* com a vocabulari de referència per a les àrees geogràfiques ha permès incorporar-ne les coordenades geogràfiques i aprofitar-ne l'estructura jeràrquica. En l'àmbit tecnològic, el passatge d'un entorn de prototipatge a un de producció no ha estat mancat d'obstacles i replantejaments. En el futur es preveu enriquir el projecte amb la descripció de les característiques ampelogràfiques, agronòmiques i enològiques de les varietats, fomentar les aportacions dels usuaris a través del web i promoure aquest projecte al núvol de dades enllaçades.

### Resumen:

*Vitis catalana* ([www.vitiscatalana.cat](http://www.vitiscatalana.cat)) es un portal especializado en las variedades de vid en los territorios de habla y cultura catalana. Ofrece información histórica y terminológica sobre estas variedades, basándose en el vaciado de documentos en forma de citas de presencia. El proyecto se ha desarrollado con una aplicación basada en el modelo de datos RDF y tomando como referentes estándares como el CIDOC CRM y SKOS. El uso de *GeoNames* como vocabulario de referencia para las áreas geográficas ha permitido incorporar las coordenadas geográficas y aprovechar su estructura jerárquica. A nivel tecnológico, el pasaje de un entorno de prototipado a uno de producción no ha sido carente de obstáculos y replanteamientos. En el futuro se prevé enriquecer el proyecto con la descripción de las características ampelográficas, agronómicas y enológicas de las variedades, fomentar las aportaciones de los usuarios a través de la web y promover este proyecto en la nube de datos enlazados.

Paraules clau: varietats de vinya, ampelografia, RDF, *Linked Open Data*, *GeoNames*, CIDOC CRM, SKOS.

Palabras clave: variedades de vid, ampelografía, RDF, *Linked Open Data*, *GeoNames*, CIDOC CRM, SKOS.



## 1. Introducció

*Vitis catalana* ([www.vitiscatalana.cat](http://www.vitiscatalana.cat)) és un portal especialitzat en les varietats de vinya als territoris de parla i cultura catalanes. Les primeres evidències de producció de vi català daten del s. I aC. L'escriptor llatí Plini el Vell<sup>1</sup> en parla, i les restes arqueològiques ho corroboren. Però, quins vins s'han consumit al llarg dels segles? On es cultivaven? A partir de quines varietats s'elaboraven? Quina és la varietat més típica d'una zona? Quines variants dialectals té una varietat? Aquestes preguntes miren de ser resoltes amb l'eina *Vitis catalana*, un recurs que pretén ajudar els historiadors i el públic en general a teixir arguments sobre la presència, al llarg del temps, de determinades varietats de ceps als Països Catalans.

El projecte de *Vitis catalana* està liderat pel Centre de Documentació VINSEUM (CDV) i està emmarcat dins del Grup de Recerca de les Varietats Autòctones i Tradicionals de la Vinya de l'àrea dels Països de parla i cultura Catalana (GREVA). Aquest grup està impulsat per VINSEUM, Museu de les Cultures del Vi de Catalunya, i la Institució Catalana d'Estudis Agraris (ICEA), filial de l'Institut d'Estudis Catalans, i hi participen persones amb una llarga trajectòria en el món de la vitivinicultura.

En aquesta comunicació mirarem de descriure la *Vitis catalana* des de dos punts de vista: el del contingut i el del continent.

## 2. El contingut: les varietats de vinya

*Vitis catalana* consta de dos blocs temàtics: el bloc històric i terminològic, i el bloc descriptiu. El primer té com a objectiu recollir, sistematitzar i presentar les cites de presència de les varietats de vinya localitzades en documents manuscrits i bibliogràfics al llarg de la història. El segon bloc –que es desenvoluparà en una segona fase– té com a objectiu oferir informació ampelogràfica i enològica de les principals varietats cultivades actualment als Països Catalans.

### 2.1. Bloc històric i terminològic

La documentació de les varietats –procés de localitzar informació de varietats de vinya en tot tipus de documents– no es pot entendre sense relacionar-la amb la història de l'ampelografia. L'ampelografia és la branca de la viticultura que s'ocupa de classificar i descriure botànicament les diferents varietats de vinya. Segons l'historiador Emili Giralt, la «conveniència d'estudiar les característiques morfològiques de les varietats de ceps per tal de distingir-les clarament entre si i superar el gran obstacle de la sinonímia va ser sentida ja en el món antic. L'agrònom hispanoromà Columel·la va estudiar-ne 58, amb

---

<sup>1</sup> Gai Plini Segon. *El llibre del vi*. Barcelona: Entrecomes, 2015.

notable precisió. La decadència general de les ciències durant l'Edat Mitjana repercutí també en el terreny ampelogràfic, que fins a l'aparició de l'obra de De Crescentiis –l'any 1304– no va rebre cap aportació interessant. L'agronomia renaixentista, amb els noms d'Alonso de Herrera, d'Agostino Gallo i d'Olivier de Serres, no mostrà gaire curiositat pel tema ampelogràfic. Per contra, durant la Il·lustració els estudis de Duhamel du Monceau, Rozier i Clemente y Rubio són peces fonamentals d'aquesta part de la ciència botànica».<sup>2</sup>

En el context català, és a partir de l'edat mitjana quan apareixen les primeres notícies de varietats, les quals acostumen a ser breus, poc o gens ampelogràfiques i solen coincidir amb un tipus de vi. És el cas d'un fragment de *Lo Crestià*, de Francesc Eiximenis, datat el 1384: «Qui posa una lletra que un gran golafre eclesiàstic tramés a un metge per demanar-li sobre el seguiment de sa vida [...]. Mon veure, quan al vin blanc, és aquest: bec grec d'estiu, e d'hivern cuit, o moscatell, malvasia, tribia, còrcec, o candia, o vernassa [...]».<sup>3</sup>

A partir de la segona meitat del s. XVIII, la necessitat econòmica –o la necessitat d'obtenir més i millor rendiment de la vinya– impulsen la necessitat d'estudiar les varietats, amb l'objectiu de poder distingir aquelles de més qualitat i més productives. A partir de la segona meitat del s. XIX, les fonts en les quals apareixen notícies sobre varietats augmenten de forma considerable, sobretot a causa de l'aparició de les malalties que ataquen la vinya catalana –i europea. L'oïdi i el míldiu impulsen la recerca de varietats més resistents a la malaltia; la fil·loxera comporta una autèntica revolució al camp, amb l'estudi de les millors combinacions entre peus americans i varietats de *Vitis vinifera*.

La documentació on es pot localitzar aquest tipus d'informació és molt variada: obres literàries, tractats o manuals, descripcions històriques i geogràfiques, diplomataris, capbreus, compravendes i permutes, contractes agraris, capítols matrimonials, testaments, inventaris *post mortem*, catàlegs d'exposicions, catàlegs i publicitat d'empreses de viviers, revistes d'agricultura, informes governamentals, llibres de viatges, etc.

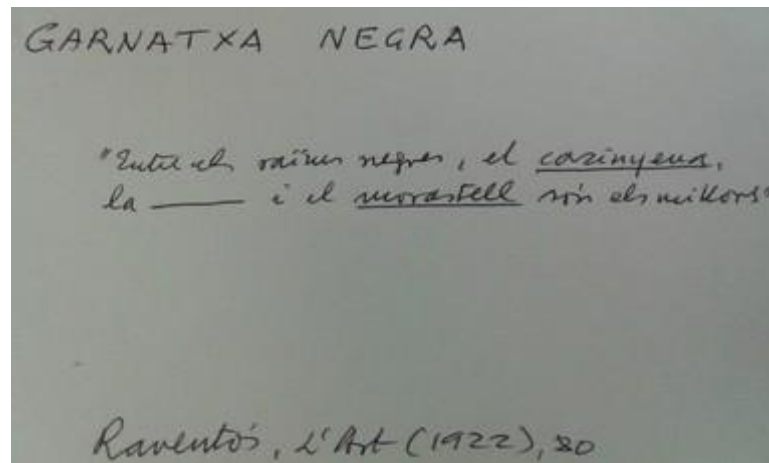
L'historiador Emili Giralt va treballar durant molts anys en la construcció del *Fitxer general de varietats de ceps*, actualment conservat al CDV. Està compost d'unes 15.000 fitxes manuscrites, ordenades alfabèticament per nom de varietat, sinònim o variant dialectal, la majoria de les quals amb presència als Països Catalans. Aquestes fitxes s'estructuren de la següent manera:

---

2 GIRALT I RAVENTÓS, Emili. «El conreu de la vinya», *Història Agrària dels Països Catalans*. Fundació Catalana per a la Recerca [etc.], vol. III (2008), p. 331-393 (p. 380).

3 EIXIMENIS, Francesc. *Lo crestià*. Barcelona: Edicions 62, 1983.

- Encapçalament principal: nom de la varietat, tal i com apareix en el document.
- Cita o comentari: fragment del document en el qual apareix la varietat (entre cometes), la zona geogràfica i/o el color de la varietat.
- Font: cita bibliogràfica d'on s'ha extret aquesta informació, i la pàgina concreta del document.



Exemple d'una fitxa del *Fitxer general de varietats de ceps*

La quantitat d'informació dipositada en aquests fitxers és molt elevada, tot i que no tota té la mateixa importància qualitativa. Les fitxes que inclouen informació descriptiva de la varietat són minoria i el nombre de varietats que inclou encara no s'ha pogut determinar, igual que la quantitat de fonts documentals utilitzades. Gairebé la majoria de fitxes van associades a una zona geogràfica, que pot ser un poble, un municipi, una província, una comunitat autònoma o un país. Sobre l'abast cronològic, en la majoria de casos s'utilitza l'any d'edició del document per situar la varietat en un punt cronològic.

## 2.2. Bloc descriptiu

Es tracta del bloc de *Vitis catalana* que descriu les característiques ampelogràfiques, agronòmiques i enològiques de les varietats. Es tracta d'un bloc eminentment textual i gràfic. De fet, la informació que conté és un resum del treball d'investigació que el GREVA està realitzant i, per tant, elaborat per agrònoms, enòlegs i sommeliers. Aquesta informació s'anirà introduint a *Vitis catalana* a mesura que s'avancin aquests treballs.

Pel que fa a les característiques ampelogràfiques, el portal ofereix informació sobre nom i sinonímies, descripció genètica, descripció dels òrgans (fulla, sarment i raïms, baia, etc.), vigor i port, i fenologia. Les característiques agronòmiques inclouen informació com la superfície conreada i distribució en les diferents DO, l'adaptació edafoclimàtica, les característiques de conreu, la sensibilitat a les malalties, els rendiments productius i la selecció clonal. Per

últim, les aptituds enològiques fan referència al potencial enològic i tipus de vins, a la composició química del vi i a la descripció sensorial.

No totes les varietats podran disposar de tota aquesta informació. En un futur es preveu que almenys les varietats treballades pel GREVA disposin d'aquests capítols. De la resta, algunes vegades només es podrà dir si es tracta d'una varietat de raïm blanc o negre.

### 3. El continent: l'eina *Vitis catalana*

L'interès actual en la recuperació de les varietats tradicionals i la utilitat de les fitxes dins del treball desenvolupat pel GREVA en va motivar la sistematització, amb l'objectiu de facilitar-ne l'estudi i difondre'n el contingut.

#### 3.1. Fases del projecte

El projecte consta de les següents fases:

##### 3.1.1. Fase d'àmbit local: definició dels requeriments

Es defineix el tipus d'informació que vol recollir l'eina, així com la seva estructura, basada en els dos blocs d'informació ja descrits: el bloc històric i terminològic i el bloc descriptiu. Des del CDV es crea una base de dades documental d'entorn local, que desenvolupa el bloc històric i terminològic. Aquesta presenta una estructura basada en dues entitats principals: les cites de presència i les fitxes bibliogràfiques. Les cites de presència contenen aquella informació relacionada amb la varietat apareguda en un document i es pot desgranar en els següents camps:

- Variant lèxica: nom de la varietat tal i com apareix en el document.
- Varietat: nom predominant i més reconegut de la varietat. L'assignació d'una variant lèxica a una varietat determinada és el pas més crític de tot el procés, fins al punt que en alguns casos pot ser una tasca impossible. En aquest darrer cas, es preveu la possibilitat de no assignar cap varietat principal. L'ús de bibliografia de referència com el llibre *Diccionari dels noms de ceps i raïms: l'ampelonímia catalana*, de Xavier Favà i l'opinió dels experts del GREVA són els principals recursos per assegurar una assignació correcta.
- Àrea geogràfica: nom del lloc on, segons la font, està cultivada la varietat. En alguns casos, el terme geogràfic és molt ampli (per exemple, Catalunya), fet que cal tenir en compte a l'hora de mostrar la informació gràficament, és a dir, sobre un mapa.
- Període cronològic: any de la presència d'aquella varietat, segons el document.
- Text de la cita: transcripció del text on apareix la varietat citada.

- Imatge: instantània de la pàgina on apareix la cita, sempre que sigui possible.
- Font de referència i pàgina: enllaç amb la fitxa bibliogràfica i detalls de la pàgina o full exacte on es pot localitzar la varietat citada.

La segona entitat principal són les fitxes bibliogràfiques. Aquestes contenen la informació habitual de descripció d'una font documental: autor/s, títol, publicació i descripció física. També inclou l'accés al document original – sempre que sigui possible– i la informació sobre la localització d'un exemplar.

Per exemple, el 1678 s'edita el llibre titulat *Cataluña ilustrada*, escrit per Esteve de Corbera. A la pàgina 61 d'aquest llibre, apareix el següent fragment: «*En Rossellon Claretos, y Moscateles. En Alella, y Mataron, y los demás lugares de la costa desde Barcelona hasta Blanes que es la antigua Laietania, Claretos, y blancos tan enteros, y vigorosos que los navegan a la Indias. En Sitjes, y Falset Malvasias. En Canbrils, y Riudoms vinos blancos de licor que llaman dulces, y otros comunes. En lo demás del campo de Tarrona lustolines, Xerelos, Verdieles, Garnachas, y otros vinos blancos tintos, y aloques. En la ribera de Lobregat cerca de Barcelona Griegos, y en otras partes otros muchos de varios nombres, sino tan generosos, suficientes para el uso, y sustento común.*»

Partint d'aquest fragment, es creen tants registres de cita de presència com variants lèxiques aparegudes («Moscateles», «Malvasias», «lustolines», «Xerelos», «Verdieles», «Garnachas», etc.). A cadascuna se li assigna una varietat principal (Moscatell, Malvasia, Xarel·lo, Garnatxa, etc.); si no es coneix la varietat, s'assigna el terme 'Varietat sense especificar'. A continuació, es detalla l'àrea geogràfica. Per exemple, en el cas de la Malvasia, en el text es descriu que es cultiva a Sitges i a Falset, per tant, calen dos registres diferents per a una mateixa variant lèxica. Pel que fa al període cronològic, el contingut no detalla cap any, per tant, es pren l'any d'edició del document. Si el text parlés d'un any anterior, es donaria prioritat a aquest. Això és especialment important en obres facsímil o estudis posteriors, ja que la font de referència és el document a partir del qual s'ha obtingut la informació, però el contingut fa referència a una obra anterior. Per últim, només caldria relacionar aquesta cita de presència amb la font de la qual s'ha obtingut la informació, que de moment només són fonts documentals, però no es descarta que es pugui ampliar a fonts orals.

En aquesta primera base de dades s'introdueixen 648 registres o cites de presència, corresponents a 127 varietats diferents. Pel que fa a la bibliografia, s'introdueixen 80 fonts documentals, de les quals 50 formen part de la col·lecció del CDV.

### 3.1.2. Fase *open*: selecció de l'eina i desenvolupament

Es creu necessari fer un salt i crear una eina més complexa i disponible al web, perquè viticultors, enòlegs i públic en general pugui consultar i aportar noves notícies o cites de presència. Es planteja una eina creada en programari no propietari, de codi obert i dades enllaçades. Es busca la col·laboració amb experts, tant en l'àmbit tecnològic com pel que fa als continguts. En aquest context, es planteja un projecte en col·laboració amb l'empresa Nubilum.

En aquesta fase s'adapten els requeriments funcionals i el model conceptual de la base de dades local a la nova aplicació i es realitza un primer prototip de l'entorn de difusió de dades. El procés d'avaluació –realitzat entre finals del 2014 i 2015– serveix per detectar alguns problemes de criteri, com són la manca de precisió geogràfica d'alguns topònims, la saturació de punts en el mapa, les diferències entre documents originals i documents que citen fonts més antigues, etc.

Un cop avaluat el prototip, es procedeix a desenvolupar els entorns de descripció i difusió de l'eina final, *Vitis catalana*, a la qual s'aplica el disseny gràfic proposat per l'Estudi Diego Feijóo. Un cop avaluada i adaptada l'eina, s'introdueixen noves fitxes manuscrites. Es fa una selecció de les fitxes, començant per les de les varietats treballades pel GREVA i les fitxes de les varietats més significatives, amb l'objectiu de difondre l'eina en àmbits documentals i enològics.

### 3.2. Entorn tecnològic

Des del punt de vista tecnològic, *Vitis catalana* s'emmarca en un projecte de desenvolupament d'una aplicació web per a la descripció de continguts pensada per treballar amb RDF i orientada a l'entorn de les dades enllaçades.

L'RDF (*Resource Description Framework*)<sup>4</sup> és un model de dades per representar informació a través de la xarxa. És una de les peces clau de l'anomenat *Linked Data*<sup>5</sup> o entorn de les dades enllaçades. Aquest paradigma estableix una sèrie de bones pràctiques per exposar, compartir i connectar fragments de dades, informació i coneixement a l'anomenada web semàntica<sup>6</sup>.

---

4 SCHREIBER, Guus; RAIMOND, Yves. «RDF 1.1 Primer» [en línia]. W3C, 24/06/2014. <[www.w3.org/TR/rdf11-primer/](http://www.w3.org/TR/rdf11-primer/)> [Consulta: 15/02/2016]

5 Christian Bizer; Tom Heath; Tim Berners-Lee. «Linked Data - the story so far». *International Journal on Semantic Web and Information Systems*, 5, 3 (2009), p. 1-22. Preprint disponible a: <http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf> [Consulta: 15/02/2016]

6 Heath, Tom. «Linked Data: Connect Distributed Data across the Web» [en línia]. <<http://linkeddata.org/>> [Consulta: 15/02/2016]

En aquest entorn, els vocabularis controlats són els pals de paller d'aquesta integració a la web de dades. Els principals vocabularis de referència per a la documentació de col·leccions es troben accessibles en obert per a la consulta i utilització (*GeoNames*<sup>7</sup>, *Getty Art & Architecture Thesaurus*<sup>8</sup>, *The Virtual International Authority File*<sup>9</sup>, *DBPedia*<sup>10</sup>, etc.).

Actualment, existeix un bon nombre d'iniciatives que incorporen tecnologies semàntiques: *Europeana*<sup>11</sup>, *ResearchSpace*<sup>12</sup> o *CultureSampo*<sup>13</sup>, per exemple, en l'àmbit del patrimoni; la *BBC*<sup>14</sup>, en la gestió i publicació de continguts del portal web; així com molts altres projectes que utilitzen o publiquen dades conforme a les bones pràctiques establertes per aquests entorns.

### 3.3. Disseny del model conceptual

S'ha realitzat una cerca exploratòria que ha permès identificar alguns recursos de referència per a la descripció de les varietats de vi o de vinya: l'*OIV descriptor list for grape varieties and Vitis species* de la International Organization of Vine and Wine (OIV)<sup>15</sup>, l'ontologia *VITIS*<sup>16</sup>, basada, en part, en

---

7 GeoNames [en línia]. <[www.geonames.org/](http://www.geonames.org/)> [Consulta: 15/02/2016]

8 «Getty Vocabularies as Linked Open Data» [en línia]. The J. Paul Getty Trust: 18/05/2015. <[www.getty.edu/research/tools/vocabularies/lod/](http://www.getty.edu/research/tools/vocabularies/lod/)> [Consulta: 15/02/2016]

9 «VIAF: The Virtual International Authority File» [en línia]. OCLC: 2010-2016. <<http://viaf.org/>> [Consulta: 15/02/2016]

10 Jens Lehmann [et al.]. «DBpedia – A Large-scale, Multilingual Knowledge Base Extracted from Wikipedia». *Semantic Web Journal*. 6, 2(2015), p. 167-195. <[www.semantic-web-journal.net/content/dbpedia-large-scale-multilingual-knowledge-base-extracted-wikipedia](http://www.semantic-web-journal.net/content/dbpedia-large-scale-multilingual-knowledge-base-extracted-wikipedia)> [Consulta: 15/02/2016]

11 Bernhard Haslhofer; Antoine Isaac. «data.europeana.eu: The Europeana Linked Open Data Pilot, Metadata Harmonization: Bridging Languages of Description». International Conference on Dublin Core and Metadata Applications, 21-23 September 2011, The Hague, Netherlands (DCMI, 2011). p. 94-104. <<http://dcevents.dublincore.org/IntConf/dc-2011/paper/view/55/14>> [Consulta: 15/02/2016]

12 ResearchSpace [en línia]. <[www.researchspace.org/](http://www.researchspace.org/)> [Consulta: 15/02/2016]

13 Eero Hyvönen [et al.]. «Culturesampo—finnish cultural heritage collections on the semantic web 2.0.». Proceedings of the 1st International Symposium on Digital humanities for Japanese Arts and Cultures (DH-JAC-2009), Ritsumeikan, 2009. doi:10.1.1.214.1515.

14 Angeletou, Sofia. «Linked Data: new ontologies website» [en línia]. BBC, 30/04/2014. <[www.bbc.co.uk/blogs/internet/entries/78d4a720-8796-30bd-830d-648de6fc9508](http://www.bbc.co.uk/blogs/internet/entries/78d4a720-8796-30bd-830d-648de6fc9508)> [Consulta: 15/02/2015].

15 «OIV descriptor list for grape varieties and Vitis species (2nd edition)» [en línia]. París: OIV, 2009. <[www.oiv.int/en/technical-standards-and-documents/description-grape-varieties/oiv-descriptor-list-grape-varieties-and-vitis-species-2nd-edition](http://www.oiv.int/en/technical-standards-and-documents/description-grape-varieties/oiv-descriptor-list-grape-varieties-and-vitis-species-2nd-edition)> [Consulta: 15/02/2016]

16 Vitis [en línia]. «Crop Ontology». <[www.croponontology.org/ontology/VITIS/Vitis](http://www.croponontology.org/ontology/VITIS/Vitis)> [Consulta: 15/02/2016]



el document anterior, la *Wine Ontology*<sup>17</sup> de la Universitat d'Stanford, utilitzada com a exemple en la documentació oficial d'OWL, la classe *Wine* de la Product Types Ontology<sup>18</sup> o el projecte *OpenWines*<sup>19</sup>. Aquests recursos resultaran particularment útils en la fase del projecte en què es desenvolupi la descripció de les característiques ampelogràfiques de les varietats.

En l'àmbit de la informàtica de la biodiversitat, el *Darwin Core*<sup>20</sup> és un dels estàndards de referència per representar les cites de presència d'un organisme biològic en un espai i temps determinats. En aquest mateix àmbit, existeixen diversos esquemes destinats al tractament dels noms i conceptes taxonòmics i de les seves sinonímies, com ara el mateix *Darwin Core* o el *Taxonomic Concept Transfer Schema*<sup>21</sup>. Projectes de l'entorn d'*Europeana* com *OpenUp!* també s'han ocupat de la modelització de la diversitat de noms comuns de plantes i animals<sup>22</sup>.

En aquest projecte s'ha optat per prendre com a referència per al disseny conceptual el CIDOC Conceptual Reference Model (CIDOC CRM)<sup>23</sup>, un estàndard de l'àmbit del patrimoni cultural que té un alt grau de granularitat i extensibilitat. Compta amb algunes extensions com el FRBRoo<sup>24</sup> o el Scientific Observation Model (CRMsci)<sup>25</sup>. Aquest model està menys estretament vinculat

---

17 Stanford University. «Wine ontology» [en línia]. W3C. <[www.w3.org/TR/owl-guide/wine.rdf](http://www.w3.org/TR/owl-guide/wine.rdf)> [Consulta: 15/02/2016]

18 «The Product Types Ontology: High-precision identifiers for product types based on Wikipedia» [en línia]. E-Business and Web Science Research Group, 14/02/2016. <[www.productontology.org/](http://www.productontology.org/)> [Consulta: 15/02/2016]

19 OpenWines [en línia]. <<http://openwines.eu/>> [Consulta: 15/02/2016].

20 John Wiecek; *et al.* «Darwin Core: An Evolving Community-Developed Biodiversity Data Standard». PLoS ONE 7, 1(2012), p. e29715. doi:10.1371/journal.pone.0029715.

21 Taxonomic Names and Concepts Interest Group. «Taxonomic Concept Transfer Schema (TCS)» [en línia]. Versió 1.01. TDWG, 2007. <[www.tdwg.org/standards/117](http://www.tdwg.org/standards/117)> [Consulta: 15/02/2016]

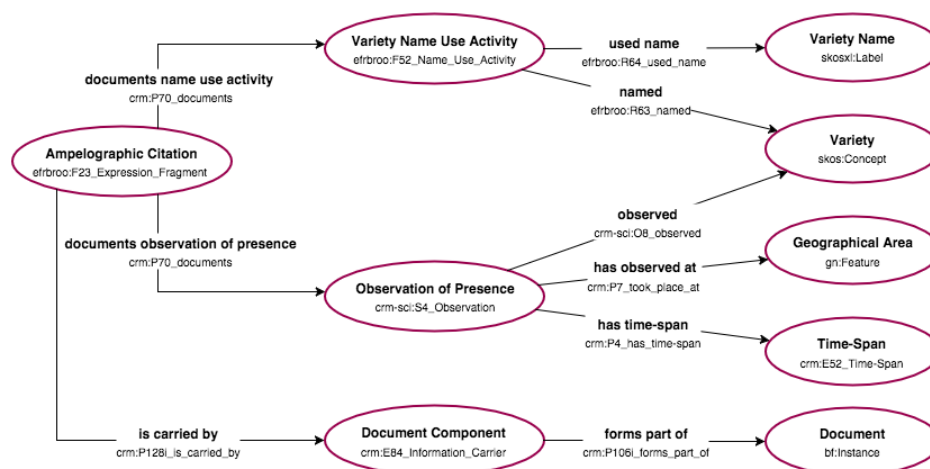
22 Rainer, Heimo; Koller, Wolfgang. «WP6 Multilingual Metadata Enrichment Including Common Names» [en línia]. Open up! <[http://open-up.eu/sites/open-up.eu/files/03.Content\\_Multilingual%20Metadata%20Enrichment%20Including%20Common%20Names.pdf](http://open-up.eu/sites/open-up.eu/files/03.Content_Multilingual%20Metadata%20Enrichment%20Including%20Common%20Names.pdf)> [Consulta: 15/02/2016]

23 Crofts, Nick [*et al.*] (eds.). «Definition of the CIDOC Conceptual Reference Model» [en línia]. Versió 5.0.4. ICOM: 2011. <[www.cidoc-crm.org/docs/cidoc\\_crm\\_version\\_5.0.4.pdf](http://www.cidoc-crm.org/docs/cidoc_crm_version_5.0.4.pdf)> [Consulta: 11/12/2015]

24 Bekiari, Chryssoula [*et al.*] (eds.). «FRBR: Object-oriented definition and mapping from FRBRER, FRAD and FRAD» [en línia]. Versió 2.4. International Working Group on FRBR and CIDOC CRM Harmonisation, 2015. <[www.cidoc-crm.org/docs/frbr\\_oo/frbr\\_docs/FRBRoo\\_V2.4.pdf](http://www.cidoc-crm.org/docs/frbr_oo/frbr_docs/FRBRoo_V2.4.pdf)> [Consulta: 15/02/2016]

25 Doerr, Martin [*et al.*]. «CRMsci: the Scientific Observation Model: An Extension of CIDOC-CRM to support scientific observation» [en línia]. Versió 1.2.2. FORTH, 2014. <[www.ics.forth.gr/isl/CRMext/CRMsci/docs/CRMsci1.2.2.pdf](http://www.ics.forth.gr/isl/CRMext/CRMsci/docs/CRMsci1.2.2.pdf)> [Consulta: 15/02/2016]

a una comunitat d'usuaris específica i ofereix un marc general d'interoperabilitat que ha de facilitar la transformació de les dades generades en l'àmbit d'aquest projecte cap a diferents esquemes de metadades, en funció de les necessitats dels diversos entorns de difusió. D'altra banda, s'han adoptat l'SKOS<sup>26</sup> i l'extensió SKOS-XL com a models de referència per a la representació del vocabulari de varietats de vinya i de les seves variants lèxiques. Pel que fa a les entitats que provenen d'altres fonts ja existents, com ara les àrees geogràfiques i els documents –aquests darrers ja catalogats en gran part al CCUC– s'han utilitzat vocabularis RDF existents com el de la *GeoNames Ontology*<sup>27</sup> i *BibFrame*<sup>28</sup>, respectivament.



Representació esquemàtica del model conceptual utilitzat a *Vitis catalana* amb les correspondències amb els models de referència.

El model conceptual s'ha formalitzat en una ontologia amb OWL. El resultat ha estat la creació d'un conjunt de classes i propietats per descriure el domini en qüestió. Totes són una especialització de les corresponents classes i propietats dels models de referència. Això permet acotar el significat i l'abast de les noves classes i propietats i establir restriccions o relacions amb vocabularis RDF que no estan definides en els models originals.

### 3.4. Dades enllaçades

Un dels objectius d'aquest projecte documental és l'elaboració d'un vocabulari raonat de noms de varietats de vinya. A banda de les obres de referència que

26 Isaac, Antoine; Summers, Ed (eds.). «SKOS Simple Knowledge Organization System Primer» [en línia]. W3C, 18/08/2009. <[www.w3.org/TR/skos-primer/](http://www.w3.org/TR/skos-primer/)> [Consulta: 15/02/2016]

27 «GeoNames Ontology» [en línia]. <[www.geonames.org/ontology/documentation.html](http://www.geonames.org/ontology/documentation.html)> [Consulta: 15/02/2016]

28 Bibliographic Framework Initiative [en línia]. Library of Congress. <[www.loc.gov/bibframe/](http://www.loc.gov/bibframe/)> [Consulta: 15/02/2016]

ja s'utilitzen actualment, com l'esmentat *Diccionari dels noms de ceps i raïms: l'ampelonímia catalana*, a mesura que avanci el projecte de recerca es valorarà l'oportunitat de vincular aquest vocabulari a altres fonts de referència, des de la *DBPedia* a l'*OIV*<sup>29</sup>, sigui a partir d'enllaços, codis o incorporant sinonímies segons les possibilitats de cada cas. Alhora, el mateix vocabulari resultant d'aquest projecte és susceptible d'acabar essent un punt de referència a la web de dades enllaçades.

D'altra banda, un dels vocabularis controlats que s'utilitza és el de descriptors geogràfics, que permet registrar les àrees geogràfiques on ha estat observada una determinada varietat de vinya. En aquest cas, en comptes de crear un vocabulari propi, s'ha optat per utilitzar directament un vocabulari extern com *GeoNames*, que està disponible amb llicència Creative Commons Reconeixement 3.0.

La utilització de vocabularis externs de referència permet, no només disposar d'una eina de normalització per a l'entrada de continguts, sinó també poder enriquir les dades mateixes amb la informació complementària que ofereixen. Les entitats geogràfiques disponibles a *GeoNames* estan ja estructurades de manera jeràrquica i disposen de coordenades geogràfiques, topònims en diverses llengües i enllaços a altres fonts d'informació, com ara la *Viquipèdia*. Això ha permès, en aquesta fase, incorporar al projecte la possibilitat de fer cerques geogràfiques jeràrquiques (p. ex. totes les varietats observades a les Illes Balears) i de mostrar els resultats de manera gràfica en un mapa.

Juntament amb la *DBpedia*, *GeoNames* és un dels nodes més ben enllaçats del núvol de dades<sup>30</sup>, és a dir, que hi ha molts projectes que disposen d'enllaços a aquests vocabularis. Aquests recursos, compartits i utilitzats per una gran comunitat d'usuaris, poden alhora servir com a eines de descoberta de les col·leccions mateixes.

L'adopció de vocabularis externs no sempre dona una resposta completa i adequada als requeriments documentals mateixos. En el cas de *GeoNames*, per exemple, les comarques no formen part de la jerarquia administrativa principal per al territori català. D'altra banda, també s'han detectat algunes mancances en les dades. En aquest segon cas, tractant-se d'un projecte col·laboratiu, s'han pogut esmenar directament amb les eines de participació

---

29 «International list of vine varieties and their synonyms» [en línia]. París: OIV, 2013. <[www.oiv.int/en/technical-standards-and-documents/description-grape-varieties/international-list-vine-varieties-and-their-synonyms](http://www.oiv.int/en/technical-standards-and-documents/description-grape-varieties/international-list-vine-varieties-and-their-synonyms)> [Consulta: 15/02/2016]

30 Cyganiak, Richard; Jentzsch, Anja. «The Linking Open Data cloud diagram». 30/08/2014. <<http://lod-cloud.net/>> [Consulta: 15/02/2016]

que ofereix la plataforma en si. Per poder tenir, doncs, un major control sobre el vocabulari, s'ha optat per crear una rèplica d'una part de les dades en un servidor local que s'actualitza periòdicament.

### 3.5. Implementació tecnològica

El passatge d'una fase de prototipatge al desenvolupament d'una eina prou estable per utilitzar-se en un entorn de producció no ha estat mancat d'obstacles i replantejaments. L'RDF i les dades enllaçades són entorns pensats originalment per a la difusió de dades. L'ús natiu d'aquests models en entorns de producció de dades no està encara prou consolidat en tots els aspectes. Tret d'algunes iniciatives com la *Linked Data Platform*<sup>31</sup>, en general s'han trobat a faltar referents tecnològics consolidats que permetin definir una arquitectura tecnològica prou estable i estandarditzada íntegrament basada en tecnologies semàntiques.

En les fases inicials, es va desenvolupar un prototip que treballava directament amb repositoris de triplets, tant per a l'entorn de descripció com de difusió. A banda de qüestions tècniques relacionades amb el control de transaccions, el consum de recursos o l'escalabilitat, algunes de les complexitats que sorgeixen són derivades de la pròpia naturalesa d'un model de dades orientat al processament per part de màquines, que s'estructura en forma de grafs i que requereix d'uns mecanismes adequats per establir els límits de cadascuna de les unitats documentals tal com les entén l'usuari d'un entorn de gestió de continguts.<sup>32</sup>

En la fase actual del projecte, s'ha optat per utilitzar un model de base de dades orientat a documents i amb una infraestructura tecnològica prou coneguda per garantir una posada en producció estable i un manteniment adequat de l'aplicació. Per treballar en aquest entorn, s'ha agafat com a referència el format JSON-LD per a la codificació de l'RDF, i s'utilitza *JSON Schema* per representar les restriccions establertes al model conceptual.

L'entorn de descripció consta d'una aplicació web SPA (*Single Page Application*) i d'una API REST amb Play Framework. Les dades s'emmagatzemen en una base de dades híbrida relacional/documental basada en PostgreSQL amb suport JSON. Per a l'entorn de difusió, les dades s'indexen en un motor de cerca *Elasticsearch* accessible mitjançant una API REST «amigable» per als entorns de desenvolupament web. D'altra banda, les dades

---

31 Speicher, Steve; Arwe, John; Malhotra, Ashok (eds.). «Linked Data Platform 1.0» [en línia]. W3C, 26/02/2015. <<http://www.w3.org/TR/ldp/>>. W3C, 26/02/2015. [Consulta: 15/02/2016]

32 Stickler, Patrick. «Concise Bounded Description» [en línia]. Nokia, W3C, 03/06/2005. <[www.w3.org/Submission/CBD/](http://www.w3.org/Submission/CBD/)> [Consulta: 10/12/2015]

en RDF són accessibles a través d'un SPARQL Endpoint amb un repositori de triplets *Apache Jena Fuseki*.

#### 4. Exemples de cerques

L'eina permet realitzar cerques simples pels següents criteris: varietat, àrea geogràfica, període cronològic i font de referència. També ofereix la possibilitat de realitzar una cerca avançada, que permet combinar diversos camps a la vegada i afegir alguna excepció. Alguns exemples de cerques avançades serien localitzar les cites de presència més antigues a Catalunya o bé les varietats més presents en un territori al llarg del segle XVIII.

L'opció simple de cerca per varietat permet recuperar totes les variants lèxiques i situar-les gràficament sobre un mapa. Per exemple, si se cerca la varietat «Moscatell», es recuperaran totes les variants lèxiques relacionades amb aquesta varietat: «moscat de taula», «moscatel», «moscatel dorado», «moscatell», etc. Entre parèntesi apareix el nombre de cites de presència de cada variant lèxica, i en fer un clic a sobre, es despleguen les cites de presència. Totes les variants lèxiques ofereixen la informació sobre la cita de presència, és a dir, de quin document s'ha extret aquella informació; i sempre que sigui possible s'ofereix el text transcrit i, si es té, una imatge del document. És l'apartat de *Vitis catalana* que emularia les fitxes manuscrites del fitxer de varietats de l'Emili Giralt. Al seu torn, les cites de presència ofereixen l'opció d'obrir la fitxa del document o font bibliogràfica, i també conèixer quines altres variants lèxiques apareixen en el mateix document. Sempre que existeixi i sigui possible, s'ofereix la ruta del document per accedir al document original.

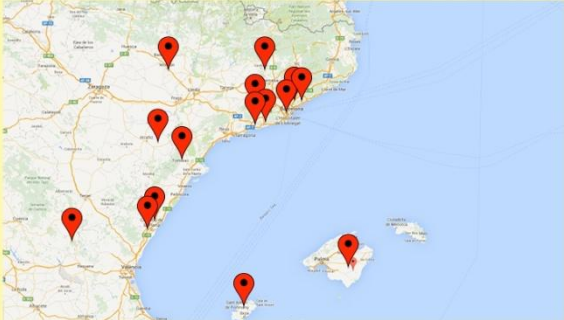
Gràcies a la pestanya «Descripció de la varietat», *Vitis catalana* ofereix l'oportunitat de conèixer els aspectes agronòmics i enològics de la varietat, sobretot si es tracta d'una de les més cultivades actualment. A través de les pestanyes «Àrea geogràfica» i «Període» es poden consultar els resultats en forma de llistat i ordenats segons un criteri geogràfic o un criteri cronològic. Els resultats també es poden exportar en format PDF.

VITIS \\\ CATALANA

---

VARIETAT  
**MoscateLL**

DESCRIPCIÓ DE LA VARIETAT
VARIANTS LÈXIQUES
ÀREA GEOGRÀFICA
PERÍODE
PDF DESCARREGAR INFORME



S'HAN OBTINGUT 18 CITES DE PRESENCIA

- Moscat de taula (2)
- Moscat negre (1)
- Moscatel (6)
- Moscatel de parra (1)
- Moscatel dorado (1)
- **MoscateLL (7)**
  - Argentona (1872)
  - Comunitat Valenciana (1935)
  - **Eivissa (1914)**
  - Piera (1893)
  - Sant Pere de Terrassa (1893)
  - Sitges (1872)
  - Ullastrell (1893)

Exemple de visualització de la pestanya de «Variants lèxiques».

Tota cita de presència té assignat un terme geogràfic. Aquest pot ser precís (un municipi, per exemple) o bé més general (un país o una comunitat autònoma, per exemple). L'eina permet cercar tant pel nivell més inferior o precís, com per un nivell més ampli, el qual també inclou els nivells inferiors. Per exemple, es poden cercar totes les varietats presents en un poble com Sant Cugat Sesgarrigues, o bé totes les varietats de la comarca de l'Alt Penedès. La ubicació en el mapa es realitza a partir de les coordenades geogràfiques associades als termes geogràfics, funcionalitat associada a la utilització de *GeoNames*.

Quan interessa localitzar les cites de presència d'una època determinada, s'utilitza la cerca per període, que permet seleccionar un any concret o bé un interval. Els resultats es mostren en format mapa i també en llistat ordenat cronològicament. Per exemple, es poden cercar totes les varietats presents abans del 1600, o bé les varietats posteriors a la invasió de la fil·loxera.

Per recuperar totes les cites de presència d'una mateixa font bibliogràfica, s'utilitza la cerca per document. La cerca es realitza pel camp autor i títol. Els resultats es mostren a través de la fitxa del document.

## 5. Conclusions

El portal *Vitis catalana* fa accessible un fons documental de gran valor per a l'estudi i el coneixement de les varietats de vinya al nostre territori. La seva missió principal és fomentar el coneixement i descoberta del patrimoni vitivinícola català. Els impulsors d'aquesta eina desitgen contribuir a la reducció del «monstruós problema –avui encara no resolt– de la sinonímia i l'homonímia ampelogràfiques» del qual es lamentava el Dr. Giralt.<sup>33</sup>

En el pla tecnològic, aquest projecte ha permès aprofundir en el coneixement dels reptes que planteja la web de dades en entorns de descripció i difusió de continguts. Cal seguir treballant per consolidar-lo i explotar de manera efectiva les possibilitats que ofereixen aquestes tecnologies. En aquest àmbit, existeixen moltes iniciatives, eines i aproximacions metodològiques diverses que cal tenir en compte per trobar l'equilibri adequat entre la innovació, la solidesa i el sentit pràctic.

Finalment, ara s'obre una fase de projecció vers la web de dades enllaçades en la qual cal establir sinergies i trobar els canals i les metodologies de difusió més adequats.

## Bibliografia

BEKIARI, Chryssoula [et al.] (eds.). «FRBR: Object-oriented definition and mapping from FRBRER, FRAD and FRSAD» [en línia]. Versió 2.4. International Working Group on FRBR and CIDOC CRM Harmonisation, 2015. <[www.cidoc-crm.org/docs/frbr\\_oo/frbr\\_docs/FRBRoo\\_V2.4.pdf](http://www.cidoc-crm.org/docs/frbr_oo/frbr_docs/FRBRoo_V2.4.pdf)> [Consulta: 15/02/2016]

BIZER, Christian; HEATH, Tom; BERNERS-LEE, Tim. «Linked Data - the story so far». *International Journal on Semantic Web and Information Systems*, 5, 3 (2009), p. 1-22. Preprint disponible a: <<http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf>> [Consulta: 15/02/2016]

CROFTS, Nick [et al.] (eds.). «Definition of the CIDOC Conceptual Reference Model» [en línia]. Versió 5.0.4. ICOM/CIDOC CRM Special Interest Group, 2011. <[www.cidoc-crm.org/docs/cidoc\\_crm\\_version\\_5.0.4.pdf](http://www.cidoc-crm.org/docs/cidoc_crm_version_5.0.4.pdf)> [Consulta: 15/02/2016]

---

33 Emili Giralt i Raventós. «El conreu de la vinya». *Història Agrària dels Països Catalans*. Fundació Catalana per a la Recerca [etc.], vol. III (2008), p. 331-393 (p. 379).



14es JCID



Col·legi Oficial  
de Bibliotecaris-  
Documentalistes  
de Catalunya

DOERR, Martin [*et al.*]. «CRMsci: the Scientific Observation Model: An Extension of CIDOC-CRM to support scientific observation» [en línia]. Versió 1.2.2. FORTH, 2014. <[www.ics.forth.gr/isl/CRMext/CRMsci/docs/CRMsci1.2.2.pdf](http://www.ics.forth.gr/isl/CRMext/CRMsci/docs/CRMsci1.2.2.pdf)> [Consulta: 15/02/2016]

GIRALT I RAVENTÓS, Emili. «El conreu de la vinya». *Història Agrària dels Països Catalans*. Fundació Catalana per a la Recerca [etc.], vol. III (2008), p. 331-393.

ISAAC, Antoine; SUMMERS, Ed (eds.). «SKOS Simple Knowledge Organization System Primer» [en línia]. W3C, 18/08/2009. <[www.w3.org/TR/skos-primer/](http://www.w3.org/TR/skos-primer/)> [Consulta: 15/02/2016]