Purdue University

Purdue e-Pubs

Flora Graeca Digitalis: Old Library Stock

Wolfgang Stille
University and State Library of Darmstadt, stille@ulb.tu-darmstadt.de

FLORA GRAECA DIGITALIS: OLD LIBRARY STOCK MEETS LINKED OPEN DATA, DIGITAL STROLLING, AND MOBILE DEVICES

Wolfgang Stille
University and State Library Darmstadt, Germany
stille@ulb.tu-darmstadt.de

Abstract

Libraries have been digitizing old stock for several years now – mostly for a specific and restricted scientific user group. But what about the social obligation to preserve, and therefore call to mind digital cultural heritage in a contemporary manner?

Flora Graeca is a large format masterpiece of printing, engraving, color and design of almost thousand plants of the Mediterranean area in the late 18th century comprising ten volumes. It was discovered in Darmstadt's University and State Library three years ago. Encouraged by the overall good condition and the finely crafted, hand-colored illustrations, the library conducted a high-quality digitization of the whole work in order to make it digitally available to the public. The key question of the work presented here was how to reach a circle beyond scientific users, i.e. botanists and historians, and draw the general public's interest to this wonderful masterpiece. The offer should address in particular young student library patrons, who use the library in hundreds as a place of work and communication. Why not making them curious in the libraries treasures while coming here every day? So, the idea was to enrich the illustrations by further information about the botanic illustrations by linked open data from the semantic web, such that entities could be linked semantically, and the digital copies could be connected to additional information. In combination with a presentation platform developed in a former project particularly for mobile usage, a virtual edition of Flora Graeca was formed, implemented as an edutainment application with responsive design and exploratory search for an intuitive and ludic handling. It breaks fresh ground by leaving the classic search and find paradigm towards digital strolling: just like a physical visit to a library or a museum, it allows strolling through a collection, discovering and comparing objects, and getting inspired.

Keywords: Digital Strolling, Linked Open Data, Digitization, Presentation, Mobile Devices.

INTRODUCTION

Scientific libraries are spots of dichotomy: lots of highly valuable special collections of old stock printed materials gently snooze in their storages and might be discovered and even utilized – if ever – only under sharp restrictions by single individuals, for scientific purpose only, of course. Growing digitization activities help only to some extent: there are comprehensive sets of digital copies yet in the libraries' digital repositories, shining stars besides ordinary born digitals, mostly accessed through a specific search query as usability and presentation levels are poor in many cases. Moreover, the possibilities of new search forms such as exploratory [Klouche et al., 2015; Kules, Capra, Banta, & Sierra, 2009; Marchionini, 2006] or semantic [Deuschel, Greppmeier, Humm, & Stille, 2014] navigation options are often not present at all. Therefore, the probability to discover digital treasures by accident is very low. Last but not least, there is the larger part of library patrons who think that the whole library is digitally available on their handheld devices (or at least should be), and any kind of content should be accessible in an easy, fast, and intuitive way. Different age brackets, different patron groups, and different subject disciplines might think thoroughly different about these theses. So far, so good. The key questions we want to pose here are 1) "What does these heterogeneous groups connect?", 2) "What can libraries itself do in order to bring those closer together?", and 3) "How can libraries encourage the young digital public's curiosity and interest in their cultural treasures?" – these are exactly the questions that challenge us in a nearly 450 years old library, that is university and state library at the same time, that is visited by around 2000 heterogeneous patrons – hobby historians, full-time scientists, and students for the most part – a day, and therefore has to face and – wherever possible – bridge these gaps on every occasion.

In this article, we show as a proof of concept how to connect historical old stock to smart devices by an edutainment application that enriches the digitized entities by linked open data from the semantic web. We use a presentation platform optimized for intuitive, exploratory, and ludic handling on mobile devices and get the *Flora Graeca Digitalis*. The utilization of semantic web resources has previously been found beneficial in digital cultural heritage applications [Bermès, 2014], in particular when there are little human resources to gather additional information that is essential for such scenario. In a similar

FRAXINUS.

Linn. G. Pl. 550. Juss. 105. Gaertn. t. 49.

TABULA 4.
FRAXINUS ORNUS.

Fig. 1. Source material of *Flora Graeca*, exemplarily the botanical description and illustration of *Fraxinus Ornus* (manna ash), the fourth of 966 hand-colored tables of the plants of Greece in the late 18th century, resulting from a survey by John Sibthorp and Ferdinand Bauer.

manner *edutainment* or *infotainment* applications do, we aim at a combination of education, information and entertainment in order to attract attention to cultural heritage content, with high demands on usability and aesthetics on mobile devices. We implemented this application as a prototype, because we believe that library old stock is an important and interesting part of cultural heritage, and thus must be presented in a contemporary way in order to be present and to be perceived by a broad and mainly digital public. Otherwise, this kind of content would usually be reserved to a very small group of people, and therefore vanish from the consciousness of the general public sooner or later.

Preliminaries

During the last years, the department of electronic information services of Darmstadt's University and State Library was involved in several projects developing a platform for cultural heritage institutions – namely libraries and museums – with the goal to expand their real physical exhibition spaces to the digital reality with a thoroughly educational claim [Eschenfelder et al., 2013]. Topics such as digital storytelling [Deuschel, Heuss, & Broomfield, 2014] and serendipity framed the scope of the project. One of the main goals was to leave the classic search and find paradigm and break new ground towards a new direction in the digital universe, that is in fact a really old one in the physical world: a real visit to a museum or a library usually offers a much wider variety of perception forms than a digital experience: we can stroll around, get inspired, be more focused, diffuse, or simply curious. Why are we apparently so content with the digital world, lacking essentially almost everything we appreciate in the physical one? This should be overcome – at least partly – by the projects' developments, precisely amongst others alternative navigation methods which allow *digital strolling* [Deuschel, Greppmeier, et al., 2014; Deuschel, Heuss, & Humm, 2014]. In the end, the digital collection of no less a figure than Frankfurt's Städel Museum was implemented by mean of the platform and its techniques, and is now online for more than one year[1]. See [Stille, 2015] for a survey on the projects' results. An ambitious idea at an

---

[1] https://digitalesammlung.staedelmuseum.de

early project stage was to use the platform in a similar way in order to present high quality digitized old stock, but due to the lack of semantic information, and the lack of human resources being able to gather and index such information, this was discarded soon.

One year later, a complete and very well-preserved exemplar of Flora Graeca by John Sibthorp was discovered by chance in the storages of the library. This work contains 976 hand-colored illustrations of plants growing in Greece and is ranked as the most costly and beautiful book devoted to any flora; the finely crafted and illustrated work was of both, scientific and horticultural interest [Harris, 2007; Lack, H. W.; Mabberley, 1998]. In close collaboration with the library's restoration department, our digitization lab digitized all ten volumes with an ultra-high (50 megapixels) resolution using a Hasselblad CFV-50C digital back that was acquired especially for the purpose of digitizing large formats at very high resolution. The digital copies have been made publicly available[2] in one of the library's repositories that underlies exactly the properties mentioned above: the digital content will be found – as the physical one – only by expert scientists, or by accident, or both. Our initial idea of an automatic indexing in order to generate further metadata was quickly withdrawn as the used fonts were not OCR compatible at all, and the describing texts were written in modern Latin, which is not a good candidate for machine translation and text mining. So, we passed the digital copies to a botanist in order to do a manual indexing at least by their old botanical names and their corresponding current botanical names according to the Linnaean biological classification as well. This gave rise to the idea that finally provides the basis to the existing application: using the botanical name or a derivative, we are able to access further information from the sematic web, for example content from DBpedia [Auer et al., 2007; Lehmann et al., 2015]. These are available under the Creative Commons-Attribution-ShareAlike-3.0 license, which allows free usage of the data under an identical license if the credits are given to the licensor. In our use case of botany, we were able to access data such as descriptive information about the plant, in English as well as German, its common name, the parameters of its scientific classification, a current photo, and – via the classification – further information about the plant's family, order, and genus. This information is used on the one hand to present further information about a single plant in our application, on the other hand to semantically link plants to each other via their scientific biological classification. Moreover, it allows a semantic browsing-like navigation through the collection by occasion, or by related objects, e.g. plants of the same family, order or genus. We call this kind of navigation *digital strolling*, as it resembles the user's motion through physical collections in the digital world, which might be quite focused on the one hand, but also very diffuse or just curious on the other hand. Thus, users are able to drift through the collection without looking for something concrete. They leave a path behind them that allows at any time to go back an arbitrary number of steps, and branch into a new path at every object. The path in turn shows the relations between the visited objects in terms of their botanical classification.

Information Harvest and Platform Ingest

As shown in Figure 1, the original work contained a Latin description of the plant. Therefore, botanists collected the old Latin botanical names solely by intellectual indexing and translated them into the well-known taxonomy of Carl Linnaeus. Only the Latin botanical name itself was too little information for a semantic application. Human resources for the intellectual collection and inquiry of further information were not available, not even in the medium run. Therefore, automatic methods of information retrieval had to be applied. We tried to gather further information from the descriptive text by OCR and machine translation which failed due to the fact that (1) machine translation of Latin is not available for productive usage at all, (2) there are too many abbreviations in the descriptions, that were not automatically resolvable, (3) the texts consist of a mixture of Latin and Greek characters that aggravate an efficient OCR. In order to enrich information ideally by automatic methods, another attempt aimed at using linked open data. We experimented with several data sources and found out quickly that DBpedia would deliver a reasonable set of features of high quality and a good coverage. Therefore, the original name from Flora Graeca, or the translated botanical name according to the Linnaean biological classification, and derivatives of it were used to send a SPARQL query to DBpedia's SPARQL endpoint getting back results via RDF triples in around 70% of the queries. In particular, the data requested from DBpedia was the common plant name, its scientific classification consisting of plant family, gender, and order, full texts of a plant description and further descriptions about the scientific classification such as a description of the plant family, and finally a URL containing a current photo of the plant. This was automated by a small piece of Python code and finally processed as ingest of the platform, whereas the data mentioned above was used as additional information, and eventually the plant family, order and genus were utilized to build a SOLR index with the purpose to semantically search and link plants to each other. In other words, we built typed networks connecting the set of objects (plants), that make it possible to move between
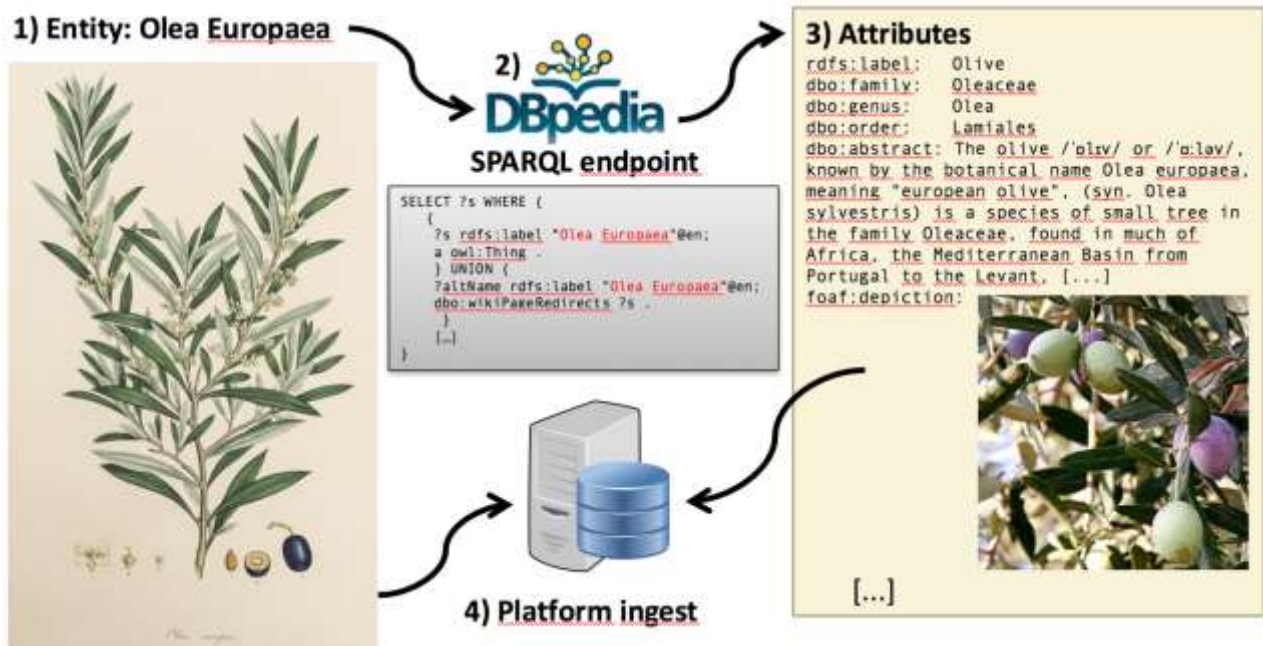
---

[2] http://tudigit.ulb.tu-darmstadt.de/

Fig. 2. Using the manually indexed entity, in this example Olea Europeae (1), a query is sent to the SPARQL endpoint of DBpedia (2) delivering attributes of the entity as RDF triples (3). These are used together with the name of the entity and the digital copy of the illustration as ingest for the platform.

two objects on one (or more) typed edge(s) showing the relations, similarities and differences between the two candidates. Figure 2 gives an overview about the harvesting and ingestion process.

Flora Graeca Digitalis

The resulting application acts as a classic client-server web application: all content and the semantic relations are delivered by a server module using a stateless REST architecture. The client is implemented exclusively in HTML5 and JavaScript which is very light-weight and works on virtually any relevant web browser. In particular, a satisfying user experience on mobile devices such as tablet computers with touch sensitive screens were an important aspect in UI design. The requirement to run on mobile devices as well as ordinary computers ended up in a responsive design frontend.

The entry screen of Flora Graeca Digitalis shows a choice of different plants and typed relations between them such as family, order, or genus (cf. Fig. 2). By touching a plant on the screen, it is scaled up, and a choice of other plants related to it by the typed relations is recommended for further browsing. Another click on the plant shows an overlay with additional information (cf. Fig. 3) such as descriptive texts describing the plant itself and its family. Additionally, the scientific classification is shown next to a current photo of the plant. A further click on the plant activates the full-screen mode (Fig. 6) which allows progressive zooming up to the highest resolution and moving the detail section by sweeping it with the finger. Closing the two overlays returns into the browsing mode again: tipping on other plants generates a history consisting of a path semantically connecting the viewed entities by typed relations (family, order, genus). Along these link paths, the user is able to move in different directions and explore the whole collection in this manner. The paths are saved and displayed, so scrolling back and forth or entering any arbitrary position on the path is possible at any time. This way of digital strolling is complemented by a classic extended search and faceting functionality (cf. Fig. 6), and a favorite box, where users are able to put in and share their favorite objects with other users. Thereby, they are able to curate their own albums. An interesting feature is the comparison mode which allows to put two arbitrary objects from the collection aside each other in order to directly show the differences and similarities between two species. Again, the shared characteristics in terms of typed relations are illustrated in the comparison. Figure 4 shows such a comparison of two different olive trees, Phillyrea latifolia and Olea Europaea. Both are related via their family (Oleaceae) and their order (Lamiales). This view is exclusively reserved to the digital edition, where entities might be arranged by a network structure, which could never be achieved by the original book that orders entities sequentially. Incidentally, this feature might be also a reason for senior users, who might be doubtful about new mobile technologies, to occupy themselves with those in order to benefit from semantic applications like these.

Fig. 2. Entry screen of Flora Graeca Digitalis showing a choice of different plants and typed relations between them. By click on a plant, it is shown in detail, and a set of other plants related to it by typed relations (family, order, or genus) is recommended for further browsing. This allows to digitally stroll through the whole collection. Along this link path the user leaves behind, he can move to an arbitrary position of the path at any time, and explore the relations between the plants on the path. Additional information of each object might be displayed as an overlay at any time. The navigation symbols on the right hand side allow direct navigation to the user's favorites, the search page, the user account if applicable, and a structured topical overview about the collection.

Conclusion and outlook

In this work, we have shown how to build a semantic application from no more than digitized old stock and linked open data from DBpedia. Our goal was to show a collection of the library's highly valuable old stock to a broad digital public with a thoroughly educational, but also enjoyable claim, and with high demands on usability and aesthetics on mobile devices. This was achieved by using a platform developed during an already completed third-party funded research project, that allows semantic linking of entities and innovative navigation techniques such as digital strolling for libraries and museums. Usually, such an intention requires a project and human resources for producing, indexing, annotating, and aligning content that can be used by the semantic application again. In our concrete scenario, due to the lack of human resources and funds, we focused on data that is already available in the semantic web and can be freely reused. Connecting the named entities to entries in DBpedia worked surprisingly well, such that human efforts could be restricted to the programming of the ingestion algorithm and adaptation of the user interface to our concrete scenario. This allowed reducing the overall project resources to a minimum (several programmer days) apart from the essential digitization and indexing process that had been already made before. The prototype was presented at the opening of the Flora Graeca exhibition in April 2016, where it found much enthusiasm of the visitors. In particular, the idea to act as a digital complementation of a physical exhibition seemed to be very attractive to us.

Next steps will be the transfer of the prototype to a productive scenario in autumn 2017 in order to present these magnificent illustrations to the global public. Further enhancements by other types of information such as open educational resources, for example videos of the plants and plant families would be nice in order to extend the application's educational aspect, e.g. to be used in the classroom. However, this again requires curative resources. Ongoing research is currently done in the areas of user understanding, recommender systems, and digital storytelling [Deuschel, Heuss, & Broomfield, 2014]

Fig. 3. Additional information about the plant and its family from DPpedia. Here, we see the European olive from the family of Oleacaea. Also shown is a current photo of the plant.

as an important variant of knowledge mediation, in particular to get young people interested in cultural heritage. Integration of social media is an important topic as well, which we support in the current implementation: sharing content with other users via social media platforms is already implemented. The integration of further linked open data sources seems obvious and will be established in the future. There is also a large interest by botanists all over the world, so linking to their applications and information sources is planned as well in the future.
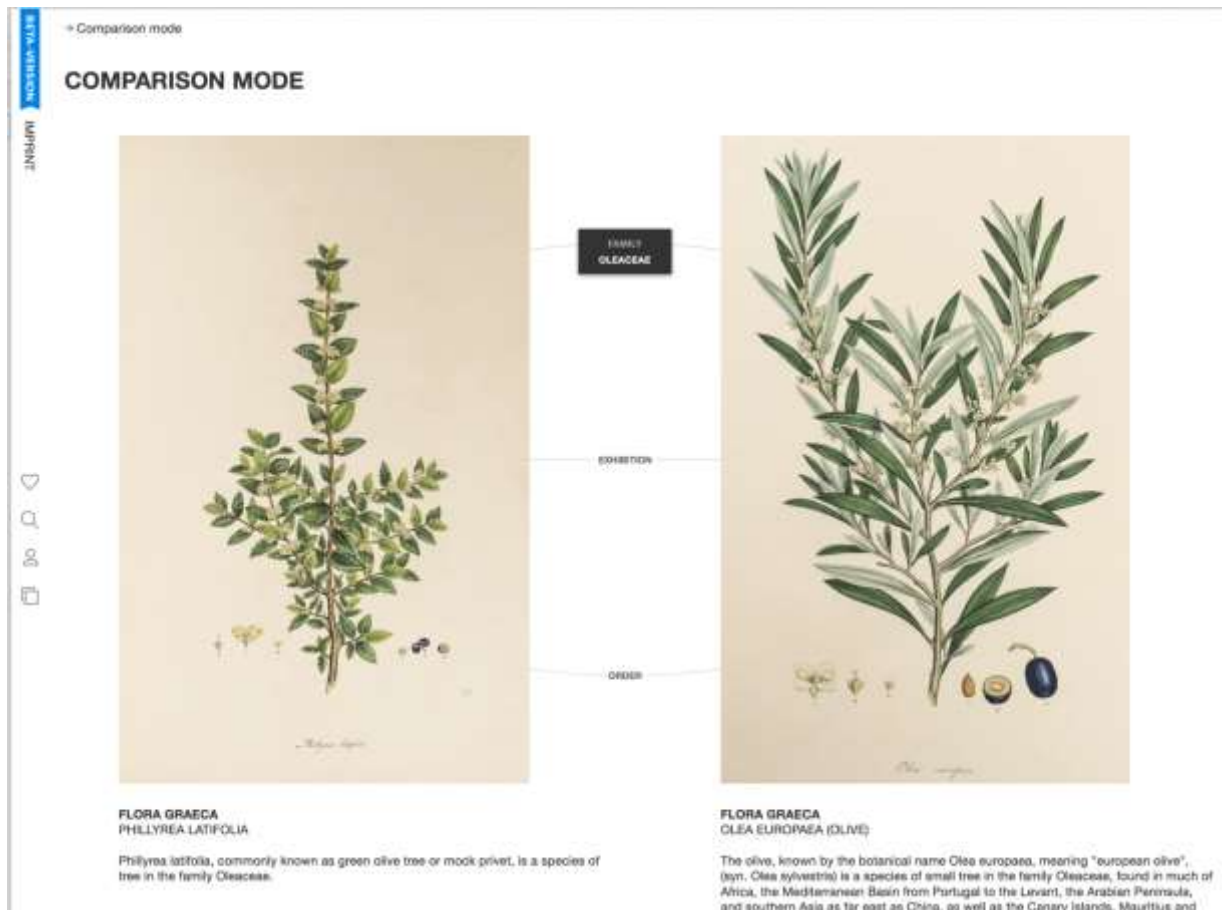
Fig. 4. Direct comparison mode: any two plants may be put aside in order to compare their details and their relations to each other. Here, two plants of the family *Oleaceae* and the order *Lamiales* are compared: the Phillyrea latifolia, ommonly known as green olive tree or mock privet on the left, and the Olea Europaea, the European olive on the right.
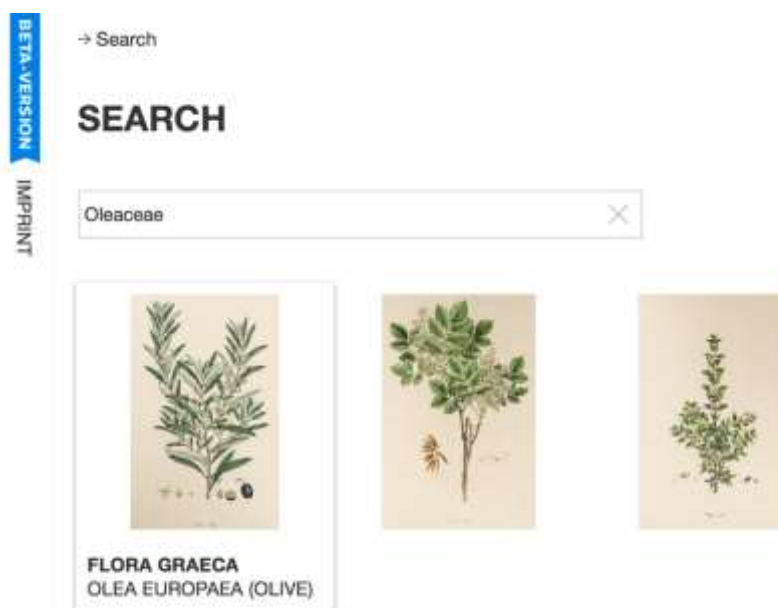

Fig. 5. Search page that show e.g. all plants of a specific family, order, or gender. Here: plants of the family Oleaceae.

Fig. 6. Detailed view: all digital copies of plant illustrations are available at very high resolutions and may be steplessly zoomed. The copy above shows the blossom of Fraxinus ornus (manna ash), a species of Fraxinus native to southern Europe and southwestern Asia.

REFERENCES

Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., & Ives, Z. (2007). DBpedia: A nucleus for a Web of open data. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 4825 LNCS, pp. 722–735).

Bermès, E. (2014). Following the user's flow in the Digital Pompidou. *Zeitschrift Für Bibliothekswesen Und Bibliographie*, *61*(2), 68–75.

Deuschel, T., Greppmeier, C., Humm, B. G., & Stille, W. (2014). Semantically faceted navigation with topic pies. In *Proceedings of the 10th International Conference on Semantic Systems - SEM '14* (pp. 132–139). New York, New York, USA: ACM Press. http://doi.org/10.1145/2660517.2660522

Deuschel, T., Heuss, T., & Broomfield, C. (2014). The Design Scope of Adaptive Storytelling in Virtual Museums. In R. Klein & P. Santos (Eds.), *Eurographics Workshop on Graphics and Cultural Heritage*. The Eurographics Association. http://doi.org/10.2312/gch.20141308

Deuschel, T., Heuss, T., & Humm, B. (2014). The Digital Online Museum: A new Approach to Experience Virtual

Heritage. In *Proceedings of the 4th International Workshop on Semantic Digital Archives (SDA 2014) co-located with the International Digital Libraries Conference (DL 2014), London, UK, September 12, 2014.* (pp. 38–48). Retrieved from http://ceur-ws.org/Vol-1306/paper5.pdf

Eschenfelder, C., Gresch, K., Fröhlich, T., Humm, B., Greiner, T., Eierdanz, P., & Blumenberg, F. (2013). The other way round: from semantic search to collaborative curation. In *Proceedings of the Nordic Digital Excellence in Museums Conference (NODEM 2013), Stockholm, Sweden, December 2013.* Retrieved from http://sched.co/HYjr6V

Harris, S. (2007). *The Magnificent Flora Graeca; How the Mediterranean Came to the English Garden.*

Klouche, K., Ruotsalo, T., Cabral, D., Andolina, S., Bellucci, A., & Jacucci, G. (2015). Designing for Exploratory Search on Touch Devices. *Proceedings of the ACM CHI'15 Conference on Human Factors in Computing Systems*, *1*, 4189–4198. Retrieved from http://dx.doi.org/10.1145/2702123.2702489

Kules, B., Capra, R., Banta, M., & Sierra, T. (2009). What do exploratory searchers look at in a faceted search interface? In F. Heath, M. L. Rice-Lively, & R. Furuta (Eds.), *Proceedings of the 9th ACM/IEEE-CS joint conference on Digital libraries* (pp. 313–322). New York and NY and USA: ACM. http://doi.org/10.1145/1555400.1555452

Lack, H. W.; Mabberley, D. (1998). *The Flora Graeca Story: Sibthorp, Bauer, and Hawkins in the Levant*. Oxford University Press.

Lehmann, J., Isele, R., Jakob, M., Jentzsch, A., Kontokostas, D., Mendes, P. N., … Bizer, C. (2015). DBpedia - A large-scale, multilingual knowledge base extracted from Wikipedia. *Semantic Web*, *6*(2), 167–195.

Marchionini, G. (2006). Exploratory search: from finding to understanding. *Communications of the ACM*, *49*(4), 41. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.97.4576&rep=rep1&type=pdf

Stille, W. (2015). Mobile Discovery for Libraries and Museums. In *Proceedings of the 36th Annual IATUL Conference*. Hannover. Retrieved from http://docs.lib.purdue.edu/iatul/2015/future/3/