

Das Museum als „Data-Warehouse“

Ein Praxisbericht über die Implementierung des Suchportals Belvedere RESEARCH ONLINE

Dagmar Diernberger – (Belvedere Wien / Research Center, Bibliothek)

Das Belvedere pflegt wie die meisten Museen seiner Größenordnung eine Vielzahl an unterschiedlichen Datenbanken zur adäquaten Verwaltung der historisch gewachsenen Sammlungen und zur Administration des Betriebs. Im Belvedere Research Center, welches im Jahr 2009 durch die räumliche und strukturelle Zusammenführung der Abteilungen Bibliothek, Archiv, Bildarchiv/Reproduktionen und Digitales Belvedere¹ entstand, werden die wissenschaftlichen Datenbanken und „Repositories“ des Hauses, die größtenteils auch über die Website abrufbar sind,² verwaltet und inhaltlich gepflegt. Um die Benutzerfreundlichkeit durch einen komfortableren Zugriff auf das vorhandene Metadatenmaterial zu steigern und die Synergien zwischen den Sub-Abteilungen weiter zu stärken, wurde 2016 das Projekt zur Implementierung des Metasuchportals *Belvedere RESEARCH ONLINE* gestartet, das die Sammlungen des Belvedere global durchsuchbar machen sollte.

Am Beginn der Überlegungen zu einem gemeinsamen Suchportal stand der Wunsch, unseren BenutzerInnen ein zeitgemäßes Interface für die Recherche nach lokalen und digitalen Ressourcen zur Verfügung zu stellen, das auch funktional allen Anforderungen gerecht werden kann. Die Entscheidung fiel letztendlich auf das cloudbasierte „Discovery & Delivery-System“ *Primo* der Firma *ExLibris*, das im Jahr 2016 neu

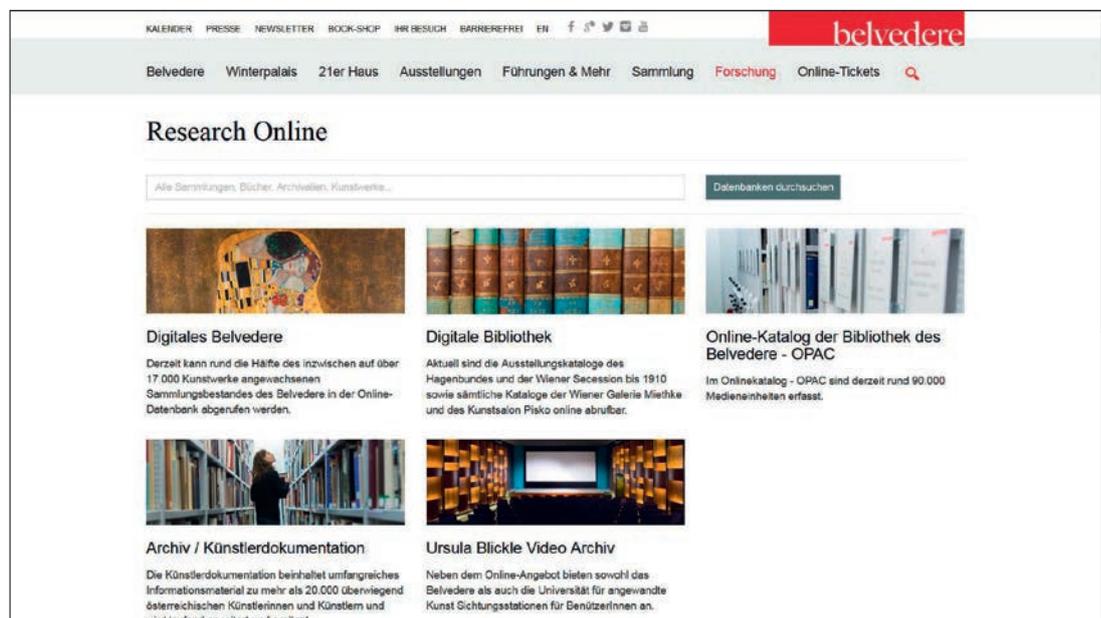
als *Primo User Interface* (kurz *Primo UI*) aufgelegt wurde. Das neue *Primo UI* kombiniert die bewährten Funktionalitäten des klassischen *Primo* Interface mit einem modernen Front-End, das sich zudem sehr flexibel an die jeweilige Corporate Identity anpassen lässt.

Im Herbst 2016 wurde mit den Implementierungsarbeiten begonnen. Ein hausinternes Projektteam – bestehend aus den Fachverantwortlichen für die jeweiligen Datenquellen und KollegInnen aus den Bereichen New Media und IT – wurde formiert.³

Zur Einbindung in der ersten Projektphase wurden fünf verschiedene Datenquellen vorgesehen, die alle über die Website des Belvedere abrufbar sind:

- Bestandsdatenbank der Bibliothek (90.000 Medien, *BIBLIOTHECAplus/OCLC*)
- Digitales Belvedere (7.000 Objekte, *eMuseum/TMS/Gallery Systems*)
- Digitale Bibliothek des Belvedere (1.300 Records, *Intranda viewer/Intranda GmbH*)
- Ursula Blickle Videoarchiv (2.600 Records, *VIMP/Iscope GmbH*)
- KünstlerInnendokumentation des Belvedere Archivs⁴ (20.000 Records, SQL Datenbank)

Abb. 1: Webauftritt Belvedere RESEARCH ONLINE.



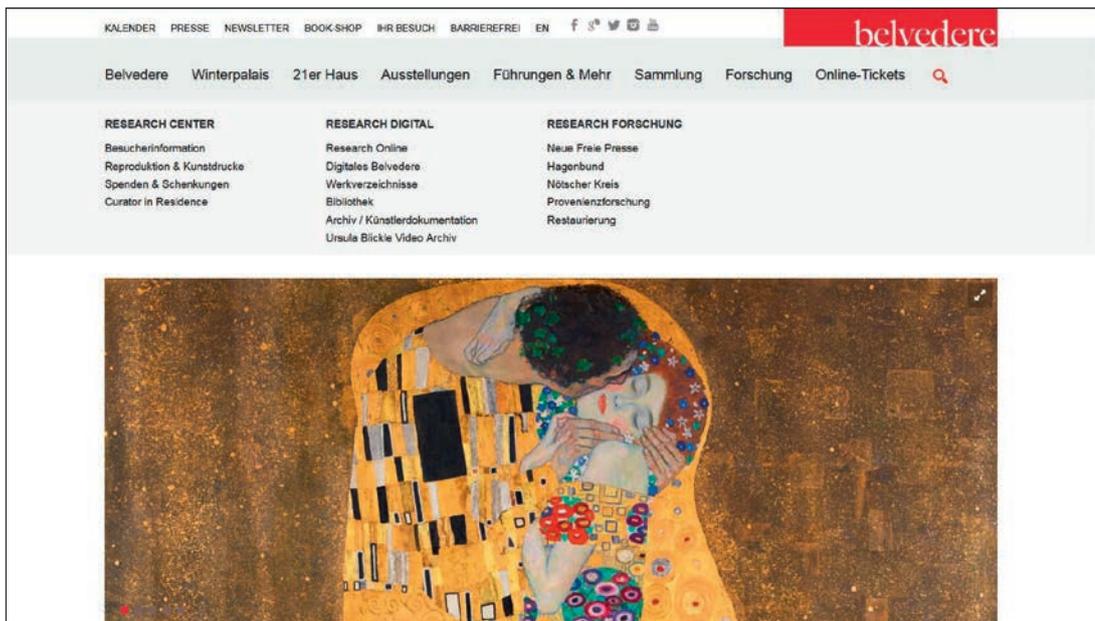


Abb. 2: Research Digital – Zugang zu den Datenbanken auf der Website des Belvedere.

Schnittstellenlösungen

Eine der großen Herausforderungen innerhalb des Projekts bestand sicherlich in der Errichtung funktionierender Schnittstellen von *Primo* zu den fünf unterschiedlichen *ExLibris*-fremden Systemen im vorgegebenen Projektzeitraum. Aufgrund der Inhomogenität der Datenquellen musste für die Datenübertragung auf unterschiedliche Schnittstellenlösungen zurückgegriffen werden.

Die beiden digitalen Repositories *Ursula Blickle Videoarchiv* und *Intranda viewer* wurden mittels „Open Archives Initiative – Protocol for Metadata Harvesting“ (kurz OAI-PMH) an *Primo* angebunden, das von den jeweiligen Softwareanbietern technisch umgesetzt wurde. Die Ausgabe der Metadaten erfolgt hierbei standardisiert und im Format Dublin Core. Die Übertragung über OAI-PMH läuft automatisiert und ist durch eine gute Interoperabilität mit *Primo* gekennzeichnet, da die aktualisierten Daten im Back-End dynamisch ins Protokoll übernommen werden.

Für die Anbindung der Datenquelle *eMuseum* konnte auf eine XML-Ausgabe zurückgegriffen werden, die bereits von der Firma *Gallery Systems* für die Datenübertragung vom Digitalen Belvedere zum *Österreichischen Kulturpool*⁵ programmiert worden war. Die Struktur der bestehenden Programmierschnittstelle (API), die bei Abruf über eine URL die Daten jeweils dynamisch erzeugt, konnte größtenteils übernommen und über eine eigene URL abrufbar gemacht werden. Diese Separierung war notwendig, um die Ausgabe bestimmter Felder etwas abändern zu können, da die für den Kulturpool programmierte Ausgabeform v. a. im Hinblick auf die Datenharmonisierung in *Primo* nicht in allen Fällen optimal war.

Datenintegration BIBLIOTHECAplus

Die Integration der Exportdaten des OCLC-Produkts *BIBLIOTHECAplus* in *Primo* stellte eine besondere Schwierigkeit dar, zumal die beiden Konkurrenzsysteme keine dynamische Schnittstellenlösung zulassen. Da für die Indizierung und die Ausgabe im *Primo*-Front-End sämtliche Metadaten zuvor in das *ExLibris*-eigene Datenformat „Primo Normalized XML“ (kurz PNX) überführt werden, war es notwendig, vor dem Gesamtexport aus *BIBLIOTHECAplus* mehrere Testläufe durchzuführen, um das ideale Ausgabeformat vor der Normalisierung zu ermitteln. Weil das in *BIBLIOTHECAplus* integrierte Exportmodul die vorhandenen Metadaten zwar u. a. in den Formaten MAB2, MARC und MARC21 ausgeben kann, der Umfang der exportierten Daten jedoch stark variiert, fiel die Entscheidung in Abstimmung mit *ExLibris* schließlich auf eine Ausgabe der Daten im Format MAB2 und gegen MARC. Dadurch wurde vor der Übergabe der Daten an *ExLibris* nun ein weiterer Arbeitsschritt nötig, mithilfe dessen die MAB2-Daten in MABXML konvertiert wurden, um sie später dann in das XML-Schema PNX übertragen zu können. Dank des Konvertierungstools „MabToMabxml“, das von der Deutschen Nationalbibliothek kostenlos zur Verfügung gestellt wird,⁶ konnte dies aber relativ rasch und unkompliziert umgesetzt werden. Während der Konvertierungsvorgänge trat jedoch noch ein anderes Problem auf, das die Darstellung von Diakritika betraf, die vom *BIBLIOTHECAplus*-Exportmodul nicht in einheitlicher Codierung ausgegeben wurden.⁷ Da auch die Umwandlung der Exportdatei in andere Codierungen mithilfe von Editorprogrammen (weder vor noch nach der Konvertierung in

MABXML und UTF-8) diesbezüglich keine zufriedenstellende Lösung brachte, musste dazu übergegangen werden, die betroffenen Diakritika zu identifizieren und vor der Konvertierung im Editor⁸ mit der Funktion „Suchen und Ersetzen“ global mit der entsprechenden korrekten MAB2-Codierung zu ersetzen. Mithilfe dieser Funktion konnten im selben Arbeitsschritt auch unerwünschte Zeichen wie das „Nichtsortierzeichen“ und das Trennzeichen für Schlagwortketten⁹, die in *Primo* nicht benötigt werden, aus den Daten entfernt werden. Durch die Einbindung dieses Vorgangs in den Workflow war nun also auch die einwandfreie Darstellung der Metadaten aus *BIBLIOTHECAplus* gewährleistet.

Datenaktualisierung

Wie bereits erwähnt, erfolgt die laufende Datenaktualisierung der beiden Datenquellen, die mittels OAI-PMH an *Primo* angebunden sind, reibungslos. Aufgrund der dynamischen Übernahme der Änderungen vom Back-End in die Schnittstelle, konnte die *Primo* Pipe für diese beiden Datenquellen für tägliches „Harvesting“ konfiguriert werden. Die korrekte Anzeige von neuen, geänderten sowie gelöschten Records ist somit zeitnah innerhalb von 24 Stunden gewährleistet.

Da die beiden Datenquellen *BIBLIOTHECAplus* und *eMuseum* aus technischen Gründen

nicht automatisiert an *Primo* angebunden werden konnten, müssen regelmäßig Exportdateien erzeugt und *ExLibris* zur Verfügung gestellt werden.¹⁰ Im Falle der Datenquelle *eMuseum* wird hierbei die aktuelle Exportdatei des Gesamtbestandes über das API abgerufen und übermittelt. Aufgrund der insgesamt größeren Datenmenge aus *BIBLIOTHECAplus* wird mithilfe des Exportmoduls lediglich die Datei mit den „ongoing extracts“¹¹ (Neuzugänge und Änderungen) geliefert und zusätzlich einmal im Jahr ein Gesamtexport durchgeführt, um auch die „deleted records“ (Löschungen) berücksichtigen zu können.

Für den Überblick über die manuellen Exporte wird eine genaue Exportdokumentation geführt.

Mapping und Datenharmonisierung

Bei der Bündelung verschiedener Datenquellen in einem gemeinsamen Suchportal und der damit auftretenden Heterogenität der vorliegenden Metadaten ist besonderes Augenmerk auf die Harmonisierung der Daten zu legen, um diese zu einem konsistenten Datenbestand zusammenführen zu können.¹²

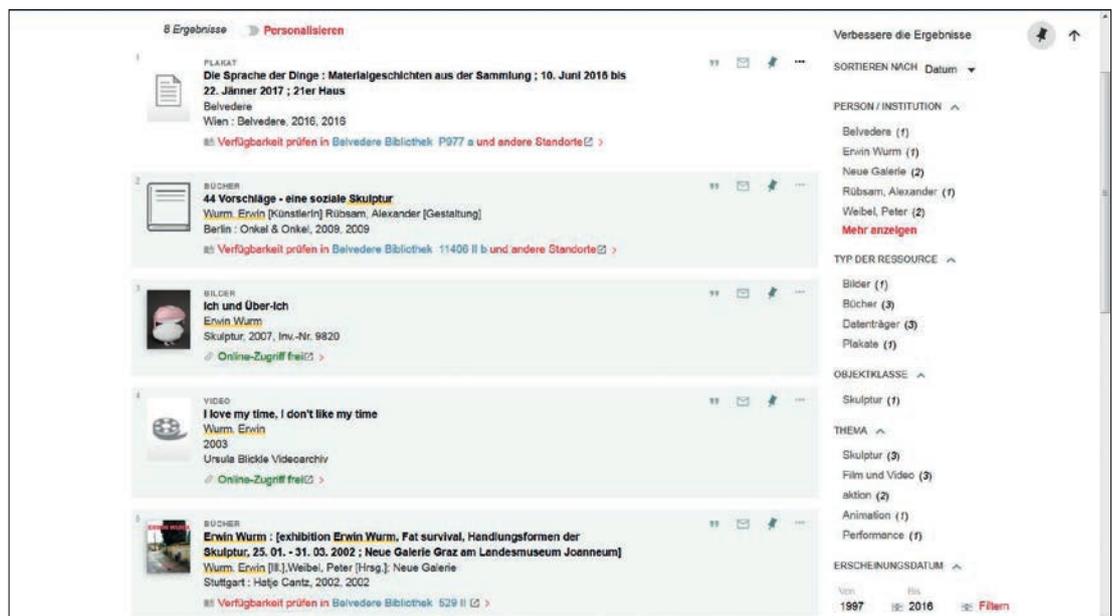
Nach der Einrichtung funktionierender Pipes zur Datenübertragung und -aktualisierung wurde als nächster Schritt in Abstimmung mit den jeweiligen Fachverantwortlichen festgelegt, welche Felder aus jeder Datenquelle in *Primo* angezeigt und durchsucht werden sollen. Zur besseren Übersicht und als Arbeitshilfe für das Datenmapping wurde eine tabellarische Konkordanz erstellt, die alle für *Primo* relevanten Feldinhalte der Datenquellen verzeichnet.¹³

Die Zuordnung der Records zu den jeweiligen Ressourcetypen wurde anhand verschiedener Merkmale in den Metadaten festgelegt. Das *Primo*-Standardset an Ressourcetypen war hier

Abb. 3: Datenharmonisierung – Äpfel und Birnen mit anderen Früchten gemeinsam in einer Schale. Fritz Schwarz-Waldegg: Obststilleben, 1927, Inv.-Nr. 6380. © Belvedere, Wien.



Abb. 4: Trefferliste mit verschiedenen Ressourcetypen.



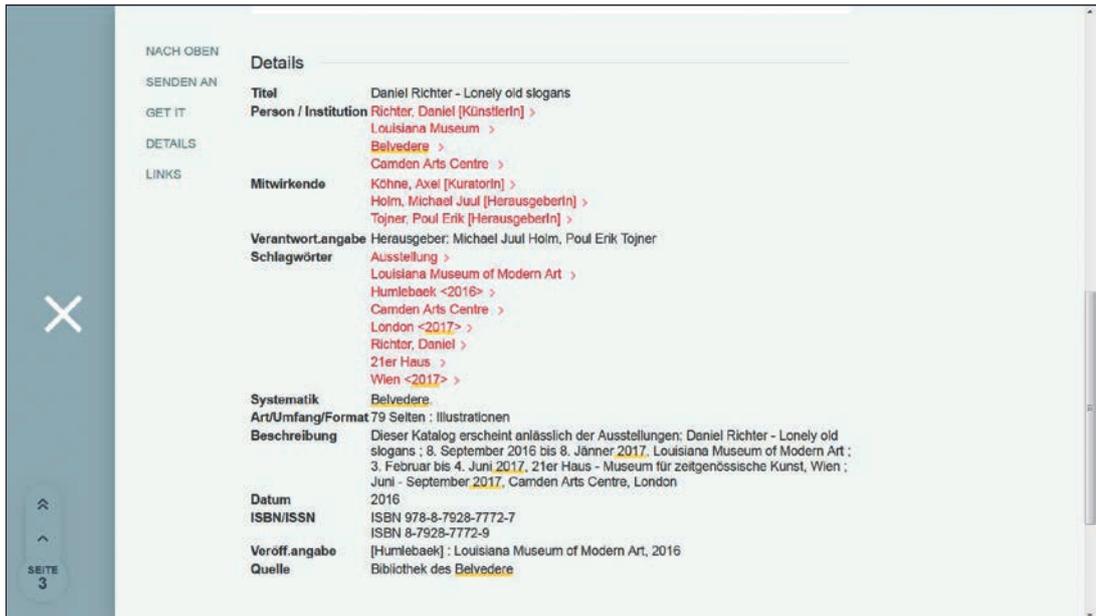


Abb. 5: Detailtreffer Bibliotheksbestand.

für den Großteil der Bestände weitgehend ausreichend; zwei zusätzliche Ressourcetypen¹⁴ wurden aber für die treffendere Beschreibung der Materialien noch erstellt.

Um eine konsistente Darstellung im *Primo*-Front-End zu gewährleisten, mussten Feldbezeichnungen gewählt werden, die für die verschiedenen Bestände Allgemeingültigkeit besitzen. Dies betraf v. a. Felder, die in mehreren Datenbanken auftreten, aber in ihrer spezifischen Bedeutung divergieren (z. B. <creator>, <subject>, <date>). Im Hinblick auf die Bibliotheksdaten wurde bei der Strukturierung der Feldinhalte¹⁵ und der Feldbenennungen¹⁶ besonderer Wert auf die Konformität der Terminologie und Struktur mit RDA gelegt, soweit dies in Abstimmung mit den anderen Datenbeständen möglich war.

Auch bei der Anzeige der „brief results“ wurde gleichermaßen Rücksicht auf alle Datenquellen genommen, da das Mapping für die Trefferliste einheitlich sein muss und die Länge der Einträge überdies beschränkt ist.

Für die vertikale Suche wurden zusätzlich vier „search scopes“ in den Suchbalken integriert, die den BenutzerInnen im ersten Suchschritt eine optionale Suche nach bestimmten Mediengruppen gestatten.¹⁷ Um die Inhalte auch facettiert durchsuchen zu können, wurden sechs Facetten gebildet, die aus Metadaten, die für alle Datenquellen relevant sind, abgeleitet wurden.¹⁸

Das neue *Primo UI* bietet überdies auch die Funktion des „FRBRisierens“ von Inhalten. Aufgrund der spezifischen Merkmale des Bibliotheksbestandes wurde diese Funktion jedoch nach einem Testlauf vorerst wieder deaktiviert.¹⁹

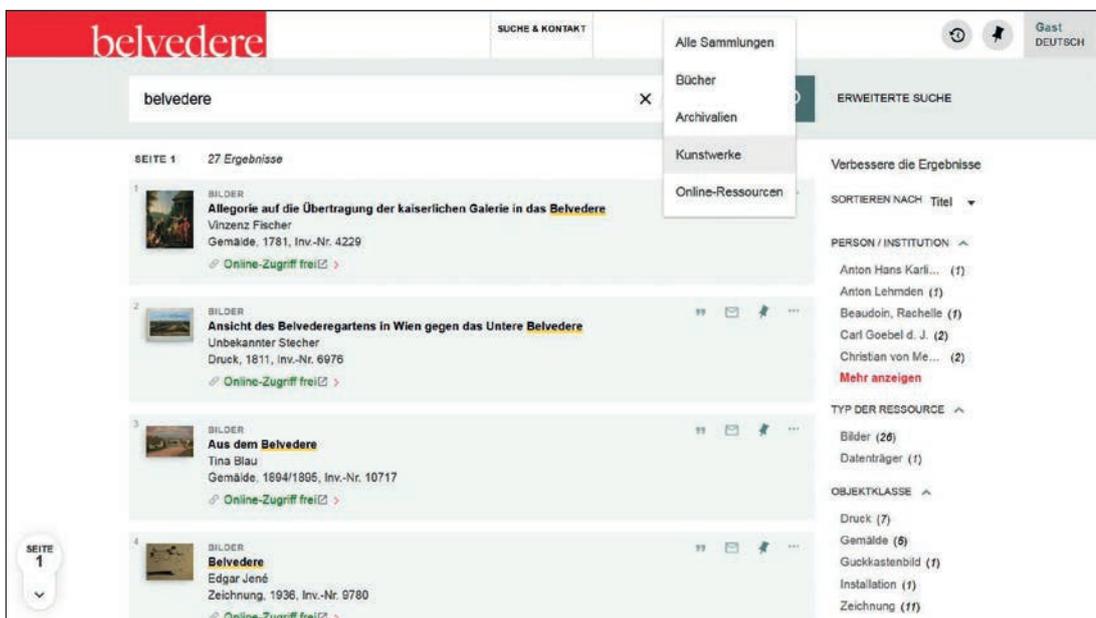


Abb. 6: Trefferselektion mithilfe der „search scopes“.

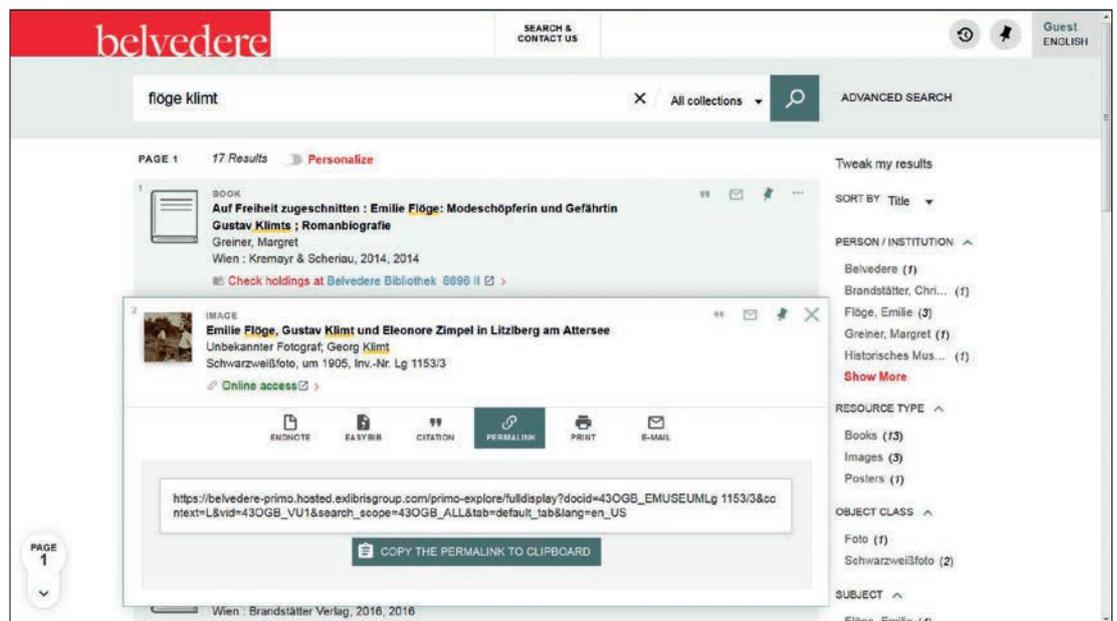


Abb. 7: Verschiedene Exportoptionen für Suchtreffer.

Funktionalitäten

Die UserInnen haben im neuen *Primo UI* also die Option sowohl zur horizontalen als auch vertikalen Suche in den Datenbeständen über die „search bar“ und zum facettierten Browsing, das eine flexible und reversible Eingrenzung der Trefferliste durch An- oder Abwählen von Facetten gestattet. Für gezielte Recherchen ist ebenso eine „Erweiterte Suche“ vorhanden. Jeder Record kann auch bereits über die Trefferliste durch eine Verlinkung in der jeweiligen Ursprungsdatenbank aufgerufen werden.²⁰ Bei der Weiterleitung zu den digitalen Repositories ist die Zugriffsregelung durch ein integriertes „Digital Rights Management“ geregelt.²¹

Für die Weiterverarbeitung der eigenen Trefferlisten sind verschiedene Exporttools²² im Suchportal fix implementiert. Auf der virtuellen Pinnwand können Suchtreffer abgelegt und ausgewählte Titel global exportiert werden. Im Hinblick auf eine internationale und interdisziplinäre Vernetzung bietet die Benutzeroberfläche überdies eine englische Sprachversion mit vollem Funktionsumfang sowie die Möglichkeit, sämtliche Inhalte über Permalinks dauerhaft elektronisch zu zitieren.

Die Anreicherung der Suchtreffer mit Thumbnails konnte für die Kunstobjekte bereits umgesetzt werden. Abgesehen von den Buchbeständen, die über *Primo* automatisiert mit Buchcovern verknüpft werden, sind auch für die anderen Datenquellen entsprechende Schnittstellenerweiterungen geplant, um die Attraktivität und Benutzerfreundlichkeit des Front-Ends weiter zu erhöhen.

CSS/HTML

Wie bereits erwähnt, lässt sich die Benutzeroberfläche des neuen *Primo UI* anhand von CSS und HTML gut an die jeweilige Corporate Identity anpassen. Die mitgelieferte „search box“²³, die der eigentlichen *Primo*-Startseite²⁴ vorgelagert wurde, ermöglicht außerdem eine optimale Einbindung des Suchportals in die Website des Museums, da die HTML-Seite abgesehen vom eingebetteten Suchbalken völlig flexibel gestaltet werden kann.

Der Vollständigkeit halber sei auch noch erwähnt, dass das neue *Primo UI* mit einem responsiven Design ausgestattet ist, welches die Darstellung bei Abruf über mobile Endgeräte automatisch anpasst.

Links²⁵

- Belvedere RESEARCH ONLINE:
<https://www.belvedere.at/de/forschung/>
- Belvedere RESEARCH ONLINE (engl. Sprachversion):
https://www.belvedere.at/bel_en/research/about_usresources
- Belvedere RESEARCH ONLINE in DBIS:
http://dbis.uni-regensburg.de/detail.php?bib_id=alle&colors=&ocolors=&lett=fs&tid=0&titel_id=102198

1. Seither wurden auch das Institut zur Erstellung von Werkverzeichnissen und das Ursula Blickle Videoarchiv im Research Center angesiedelt.
2. <https://www.belvedere.at/de> > Forschung [letzter Zugriff: 04.08.2017].
3. Primo-Projektteam: Bettina Bosin, Katinka Gratzer-Baumgärtner, Oliver Khafagi, Ralph Knickmeier, Claudia Slanar, Katharina Wallerits, Norbert Zwolinski; Primo-Projektleiterin: Dagmar Diernberger.
4. Anmerkung: An der Einbindung dieser fünften Datenquelle wird im Moment noch gearbeitet. Auch hier ist eine Anbindung mittels OAI-PMH geplant.
5. <http://www.kulturpool.at/display/kupo/Home> [letzter Zugriff: 04.08.2017].
6. siehe <http://www.dnb.de/DE/Standardisierung/Format/MAB/mabxml.html> und <https://sourceforge.net/projects/dnb-conv-tools/files/conv-env/4.1.1/> [letzter Zugriff jeweils: 04.08.2017].
7. Unterschiedlich ausgegeben wurden v. a. mit Gravis und Akut versehene Diakritika, z. B. „Á, „á“ oder „è“ [„00C1“, „00E1“ bzw. „00E8“ nach ISO 8859-1]. Ein Fehlerbericht wurde bereits an OCLC weitergeleitet.
8. Hierbei wurde vorwiegend mit dem Editor Notepad++ gearbeitet: <https://notepad-plus-plus.org/> [letzter Zugriff: 04.08.2017].
9. „-“ und „11“ [„00AC“ bzw. „00A4“ nach ISO 8859-1].
10. In Absprache mit ExLibris wurde entschieden, monatliche Aktualisierungen durchzuführen.
11. Diese Datei kann im Exportmodul mit dem Befehl „Bearbeitungsdatum ≥ TT.MM.JAHR“ erzeugt werden.
12. Sowohl die Normalisierung und Überführung aller Daten in das Format PNX, die nach dem Harvesting stattfinden, als auch die Indizierung gewährleisten dies im Hinblick auf die Durchsuchbarkeit bereits im Hintergrund.
13. Diese verzeichnet: Benennung der Inhalte im Front-End (Deutsch und Englisch), Feldbezeichnungen in den Metadaten, Feldnamen in PNX und Anzeigenreihenfolge. Zusätzlich können in der Tabelle bestimmte JA/NEIN-Merkmale mithilfe von Farbigkeit visualisiert werden (z. B. Durchsuchbarkeit, Anzeige in den Facetten, Sichtbarkeit der Felder im Kurztreffer etc.).
14. „Plakat“ und „Archivalie“.
15. Z. B.: Bündelung v. MAB100-MAB116 u. MAB200 im Feld „Person/Institution“ oder MAB410, MAB412 u. MAB425 im Feld „Veröffentlichungsangabe“.
16. Hierbei wurde v. a. die Terminologie der Suchmaschine des Österreichischen Bibliothekenverbundes berücksichtigt, die sowohl deutsche als auch englische Feldbezeichnungen enthält: http://search.obvsg.at/primo_library/libweb/action/search.do?vid=ACC [letzter Zugriff: 04.08.2017].
17. Da es bei bestimmten Medientypen Überschneidungen in den Datenquellen gibt, wurden folgende „scopes“ mithilfe eines eigenen Mappings festgelegt: „Bücher“, „Archivalien“, „Kunstwerke“ und „Online-Ressourcen“.
18. Das sind: „Person/Institution“, „Ressourcetyp“, „Thema“, „Erscheinungsdatum“, „Sammlung“ und „Sprache“.
19. Gründe hierfür sind, dass viele gleich betitelte Werke v. a. in den Mediengruppen KünstlerInnenmonografien und Ausstellungskataloge ungewollt in einem Treffer „verschwinden“ sowie die bessere Sichtbarkeit verschiedener Auflagen ohne FRBRisierung (Stichwort Museumsbestand!).
20. Für erweiterte Objektinformationen oder – falls vorhanden – zum Abruf der digitalen Ressource.
21. Digitale Ressourcen können so für UserInnen, die sich außerhalb des Heimnetzwerkes befinden, gesperrt werden. Die Information für die Zugriffsregelung wird aus dem Back-End generiert und über die jeweilige Schnittstelle mitgeliefert.
22. Das sind: E-Mail-Versand, Druckfunktion, Weiterverarbeitung durch die Literaturverwaltungsprogramme Endnote und Easybib und Ausgabe als Zitiertitel.
23. Siehe <https://www.belvedere.at/de/forschung/> [letzter Zugriff: 04.08.2017].
24. Siehe https://belvedere-primo.hosted.exlibrisgroup.com/primo-explore/search?vid=43OGB_VU1&lang=de_DE&sortby=rank [letzter Zugriff: 04.08.2017].
25. Letzter Zugriff jeweils: 04.08.2017.