

Der neue Art Discovery Group Catalogue – Anmerkungen zum Paradigmenwechsel im internationalen Kooperationsnetzwerk der Kunst- und Museumsbibliotheken¹

Jan Simone – (Kunsthistorisches Institut in Florenz)

Der 1. Mai 2014 war für das internationale Komitee des Kunstbibliothekennetzwerks artlibraries.net (ALN) ein wichtiger Tag. Nach etwa zweijähriger Vorbereitungsphase konnte ein auf dem WorldCat aufbauendes, disziplinorientiertes Discovery-Tool mit dem Namen ‚Art Discovery Group Catalogue² vor dem Publikum der Jahreskonferenz der Art Libraries Society of North America (ARLIS/NA) in Washington D.C. online gehen.³ Über die Vorgeschichte und den Kontext zur ‚Krise‘ bezüglich der Definition und konkreten Realisierung einer zeitgemäßen Fachbibliografie für Kunstwissenschaft und Kunstgeschichte ist mehrfach in mündlicher und schriftlicher Form berichtet worden, hier genügt also ein Hinweis auf entsprechende Quellen.⁴ Während die ersten Diskussionen nach 2010, dem Jahr der vorübergehenden Produktionseinstellung der International Bibliography of Art (IBA), hauptsächlich dem Ausloten verschiedener Konzepte bei der Entwicklung eines neuen und zukunftsorientierten Modells für eine ‚Kunstbibliografie‘ im weitesten Sinne mehr in theoretischer, politischer und auch finanzieller Hinsicht gewidmet waren, wurde im Frühjahr 2012 der Schritt in eine operativ-konkrete Richtung gewagt. Mit einer Pioniergruppe interessierter Partner sollten in Kooperation mit dem Betreiber des WorldCat, dem „Online Computer Library Center“ (OCLC) und seiner Zentrale für den europäischen Raum in Leiden, ein Gruppenkatalog konstituiert und die dem WorldCat immanenten Discovery-Funktionalitäten im gefilterten Sektor ausgewählter Kunstbibliotheken getestet werden.⁵ Die Projektidee wurde zügig umgesetzt und sowohl seitens OCLC als auch fördernder Institutionen wie dem Getty Research Institute und der Samuel H. Kress Foundation unterstützt. Aber auch das unerwartet hohe Interesse an einer Beteiligung an dieser Initiative hat die erste Phase entscheidend vorangebracht. Einem ‚Call for Participation‘ (August 2013) folgten innerhalb von wenigen Wochen rund 30 Kunst- und Museumsbibliotheken, und als am 1.5.2014 die

Freigabe erfolgte, waren es rund 60 Bibliotheken (zum Teil in Verbänden) aus zwölf Ländern und vier Kontinenten, deren Katalogdaten entweder bereits im neuen Gruppenkatalog sichtbar waren oder für einen Upload vorbereitet wurden. Mit einer derart guten Resonanz kann das Projektteam natürlich mehr als zufrieden sein. Ebenso gelang es, unerwartet rasch einige technische und vor allem auch finanzielle Herausforderungen zu meistern. Darüber hinaus hat das neuartige Konzept des disziplinorientierten Zugangs zum WorldCat das Interesse unserer OCLC-Partner geweckt, die nunmehr von einer „strategic collaboration between OCLC and the world’s most renowned art libraries“ sprechen und ihre Bereitschaft zur aktiven Unterstützung wiederholt zugesichert haben.⁶ Eric van Lubeek, Vizepräsident und Managing Director von OCLC EMEA⁷ und APAC⁸, bezeichnete in einem Schreiben an das Projektteam den Art Discovery Group Catalogue sogar als ein „Highlight“ vor dem Hintergrund der Mission von OCLC, „to share, expose and magnify library collections“. Können wir uns also angesichts einer so positiven Bilanz und so viel Zuspruchs seitens eines Global Players, des weltgrößten Bibliotheksservicezentrums OCLC, zurück-



lehnen und das angenehme Gefühl genießen, am Ziel angekommen zu sein? Gewiss sind wir im Erkunden von alternativen Formaten zu traditionellen Fachbibliografien ein gutes Stück vorangekommen. Doch der gegenwärtige Stand des Projekts kann bestenfalls als ein Etappensieg bezeichnet werden, die größten Herausforderungen stehen noch bevor. Diese basieren vor allem auf der neuen Erfahrung, die Bibliotheken bei der Verarbeitung und Einbindung ihrer Katalogdaten im WorldCat machen, insbesondere solche, die nicht aus dem anglo-amerikanischen Raum stammen und zum ersten Mal den Mega-Aggregator von innen kennenlernen.

Matching, Merging und Clustering – die automatisierte Zusammenführung von Daten

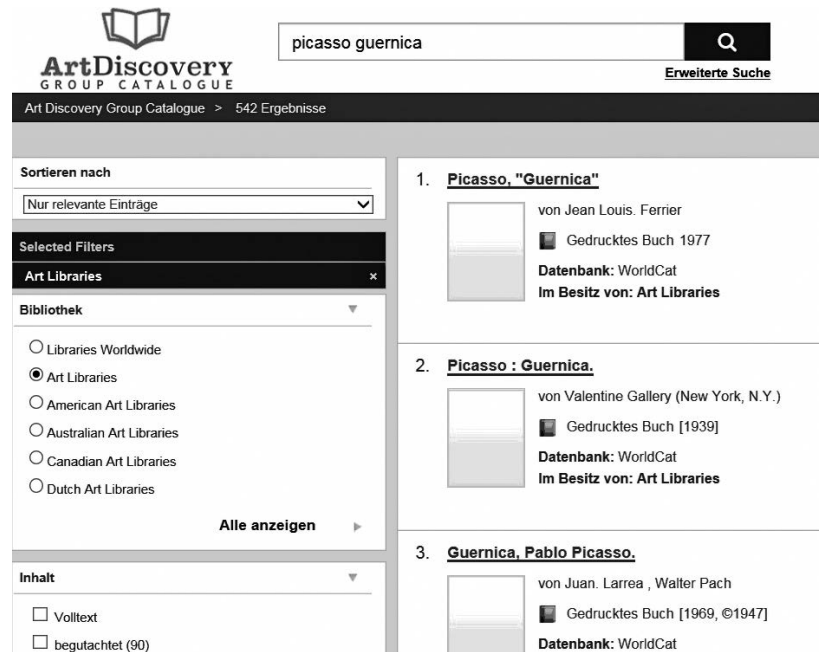
Was bedeutet es nun konkret, im WorldCat angekommen zu sein? Der standardisierte Prozess des Matching, Merging und neuerdings auch Clustering neu in den WorldCat geladener Daten führt zwangsläufig zu ihrer Veränderung, und die einzelnen Bestandteile von Katalogaufnahmen können im Einzelfall im Zusammenspiel mit analogen Dateninformationen aus anderen Quellen eine andere Funktion haben und einen anderen Informationsinput liefern als im Originalkontext. Zum besseren Verständnis sei daran erinnert, dass ein WorldCat-Gruppenkatalog lediglich eine gefilterte Sicht auf den WorldCat darstellt, die Verarbeitung der Daten also auf der Ebene des gesamten WorldCat-Bestands erfolgt, aus dem eine anhand von Besitzkennungen generierte Untermenge im Gruppenkatalog angezeigt wird. Art Discovery Group Catalogue ist folglich keine eigene Datenbank für die Kataloge der teilnehmenden Bibliotheken, sondern deren Selektion aus dem Gesamtbestand des WorldCat für eine nur scheinbar separate Ansicht- und Suchumgebung. Die Katalogdaten werden im WorldCat aber nicht einfach aggregiert, sondern nach Möglichkeit zusammengeführt, es findet eine Art Verschmelzung (Merging) statt, bei der ganze Datensätze beim Ladeprozess eliminiert werden können (Resolution), sofern ihre Übereinstimmung mit einem bereits vorhandenen ‚Master record‘ erkannt wurde und sie keine zusätzlichen Informationen für vorab definierte bibliografische Kategorien liefern.⁹ Dies kommt dem Projektziel, zuvorderst eine möglichst effiziente bibliografische Informationsressource für die Kunstwissenschaft zu generieren, durchaus entgegen. Bereits das Konzept des Netzwerks artlibraries.net zielte auf eine virtuelle Kompilation von fachspezifischen bibliografischen Informationen aus heterogenen Quellen, nicht

auf die Akkumulation von Titeldaten in einem Verbund- oder Gesamtkatalog im Sinne eines Portals zu den Beständen der Teilnehmer. Es steht bei dem Discovery-Projekt aber auch nicht die Idee einer kooperativen Katalogisierung im Raum, obgleich der sehr hohe Grad an redundanten Daten im Sektor der Kunstbibliotheken nahelegt, dass ein solches Verfahren ein überaus lohnendes Unterfangen wäre. Das Art-Discovery-Projekt dreht sich jedoch nicht um Fragen der Arbeitsökonomie, sondern es entwickelt ein Modell für eine virtuelle Fachbibliografie, basierend auf den zusammengeführten Katalogdaten ausgewählter Kunst- und Museumsbibliotheken. Die automatisierte Eliminierung (oder zumindest Reduzierung) redundanter Daten im WorldCat ist also zu begrüßen. Die Zusammenführung der Kataloge kann nun auf der operativen Basis des WorldCat auf zweierlei Weise erreicht werden: erstens, indem neue, bislang noch fehlende bibliografische Datensätze in den WorldCat geladen werden und der Gesamtbestand des WorldCat selbst und damit jener des Gruppenkatalogs bereichert wird. Und zweitens, wenn komplementäre Informationen aus geladenen Titelaufnahmen zu bereits existierenden WorldCat-Records hinzugefügt werden. Bei der letzteren Prozedur kommt es sehr auf die für den Ladeprozess entwickelten Parameter an, und die bisherigen Erfahrungen haben gezeigt, dass es sich um einen überaus komplexen und alles andere als trivialen Vorgang handelt.¹⁰ Für den Erfolg des Art-Discovery-Projekts ist es wichtig, dass analog zur bisherigen Strategie von artlibraries.net eine gleichermaßen intensive, multinationale Partnergewinnung unter Kunstbibliotheken stattfinden wird, um einen thematisch breiten, auf Komplementarität ausgerichteten, vielfältigen Datenbestand zu generieren und vor allem neue, bislang fehlende bibliografische Daten dem Gruppenkatalog hinzuzufügen. Mit dieser Strategie wird der erstgenannte Effekt – das Laden neuer Daten – erreicht. Freilich muss man sich nach den ersten Erfahrungen eingestehen, dass die Anreicherung des Gruppenbestandes mit fehlenden, unikaten Titelsätzen eher die Ausnahme sein und dass in den meisten Fällen die zuvor genannte Matching- und Merging-Prozedur die Regel bilden wird. Mit anderen Worten: Es sind Algorithmen und automatisierte Batchprozesse, die die Katalogisate der Gruppenmitglieder sowohl für die Anzeige als auch für die Indexierung aufbereiten und die Wertigkeit jedes einzelnen Datensatzes entsprechend dem Informationsinput evaluieren. Es gibt folglich Sieger und Verlierer, und die von den Gruppenmitgliedern gelieferten Titelaufnahmen müssen keineswegs zu den

Gewinnern zählen. Sie können den Gewinnern aber wertvolle Komplemente überlassen, und darauf kommt es an. Die Interaktion der Daten im WorldCat ist also die wesentliche neue ontologische Erfahrung und zugleich der prinzipielle Unterschied zu Gesamtkatalogen und Meta-Verbänden wie artlibraries.net: Die Daten können im WorldCat bzw. im Art Discovery Group Catalogue eine gänzlich andere Identität und Funktion haben als auf lokaler Ebene.

Der Mengenfaktor

Angesichts der Zielrichtung des Discovery-Projekts, ein alternatives und innovatives Modell zur traditionellen Bibliografie zu entwickeln, ist die Kenntnis der Prozesse der Datenverarbeitung im WorldCat von essenzieller Bedeutung.¹¹ Dabei sei daran erinnert, dass das zentrale Anliegen des Projekts die Bereitstellung eines Suchinstruments zur Entdeckung von dem Nutzer bislang nicht bekannter Informationen und Ressourcen ist und dass für diesen Prozess die Daten mit angemessenen, vor allem inhaltlichen Indikatoren ausgestattet sein müssen. Gewiss, eine solche Forderung trifft mehr oder weniger auf alle Kataloge zu, doch gewinnt sie im Hinblick auf den WorldCat eine besondere Relevanz, da hier nicht nur enorme Datenmengen aus höchst heterogenen Quellen aggregiert, sondern, wie dargelegt, diese entsprechend ihres Informationsgehalts evaluiert und unterschiedlich verwertet werden. Angesichts des Mengenfaktors müssen effiziente Filtermechanismen für die bei derartigen Suchstrategien unvermeidlich langen Ergebnislisten zur Verfügung stehen, die für einen disziplinierten Zugriff auf die WorldCat-Daten angemessen leistungstark sind. Sie müssen verlässlich und sollten – im Prinzip – von konstanter, guter Qualität sein, und dies hängt wiederum unmittelbar von der Qualität der zugrunde liegenden Daten selbst ab. Ein weiterer wesentlicher Faktor ist das Such- und Navigationsverhalten der Nutzer von Discovery-Systemen. Diese Rahmenbedingungen muss man sich in ihrem Zusammenspiel vor Augen halten, wenn man die Identität der eigenen Daten im Kontext des Datenmanagements im WorldCat begreifen möchte und – als mögliche Konsequenz – bereit ist, selbstkritisch die korrelierenden Standards und Traditionen der Katalogisierungspraxis zu überdenken. Letzteres ist freilich nicht das primäre Anliegen des Projekts und kann schwerlich eingefordert werden, so naheliegend es auch erscheinen mag. Doch was hat sich nun Entscheidendes für den Nutzer getan, nachdem ihm der neue Gruppenkatalog – zugegebenermaßen noch im Aufbauprozess befindlich – angeboten wird?



The screenshot shows the Art Discovery Group Catalogue search results for 'picasso guernica'. The interface includes a search bar with the query, a search button, and a results count of 542. Below the search bar, there are sorting options and a list of selected filters, including 'Art Libraries'. The main results area displays three entries:

- 1. Picasso, "Guernica"**
von Jean Louis. Ferrier
Gedrucktes Buch 1977
Datenbank: WorldCat
Im Besitz von: Art Libraries
- 2. Picasso : Guernica.**
von Valentine Gallery (New York, N.Y.)
Gedrucktes Buch [1939]
Datenbank: WorldCat
Im Besitz von: Art Libraries
- 3. Guernica, Pablo Picasso.**
von Juan. Larrea, Walter Pach
Gedrucktes Buch [1969, ©1947]
Datenbank: WorldCat

Recherchequalität – Es kommt auf die Daten an

Aus Nutzersicht vollzog sich mit der Gruppierung von Kunst- und Museumsbibliotheken im Art Discovery Tool eine ‚Googlesierung‘ des Such- und Antwortprozesses. Charakteristische Merkmale sind das voreingestellte Single-Search-Eingabefeld, die nach Relevanz gerankten Ergebnisse sowie die Facetten zur formalen und inhaltlichen Filterung. Sowohl die Algorithmen für die Relevanzrangfolge wie auch die Facetten dienen der Beherrschung der fallweise sehr langen Ergebnislisten. OCLC hat die entsprechenden Instrumentarien nach festgelegten Standards entwickelt, die im Wesentlichen den gängigen, vor allem bibliografischen Suchmaschinen entsprechen.¹² Auch im WorldCat basiert die Rangfestlegung der Relevanz auf der statistischen Auswertung der Häufigkeit der Suchbegriffe und deren Position im Datensatz. Hinzu kommen weitere Kriterien wie die Nähe der Suchbegriffe zueinander, die Aktualität der Publikation und Bestandsinformationen. Vereinfacht gesagt ist die Wahrscheinlichkeit, dass Treffer auf einer oberen Rangposition landen, umso höher, je öfter die Suchbegriffe im Datensatz vorkommen, insbesondere in den bevorzugten MARC-Feldern für Verfasser- und Titelangaben. Das Ergebnis kann jedoch erheblich von zufälligen Faktoren beeinflusst sein. Zusatzinformationen wie Inhaltsverzeichnisse, Abstracts und dgl. können die fraglichen Begriffe oft wiederholen und dem Titel eine deutlich höhere Relevanz beimessen als anderen Titeln ohne derartige Anreicherungen, obwohl Letztere faktisch eine adäquate oder sogar höhere Relevanz haben können. Auch und gerade Inhaltsdeskriptoren,

an erster Stelle Schlagwörter, spielen hierbei eine wichtige Rolle. Das zuvor erwähnte Konzept der Anreicherung von Master-Records mit ergänzenden Informationen aus geladenen Datensätzen (Merging) betrifft Daten aus den 600er-MARC-Feldern, die Inhaltsdeskriptoren enthalten, in besonderem Maße und hat somit Einfluss auf das Relevanz-Ranking. Ein solches ‚Clustern‘ von inhaltsbeschreibenden Daten, unabhängig von Herkunft und Sprache, korrespondiert mit dem Werk-Modell nach den FRBR-Standards. OCLC setzt diese Idee im GLIMIR-Projekt¹³ um und neuerdings auch im Linked-Data-Kontext, bei der Bildung der ‚WorldCat Work‘-Cluster, in denen die Inhaltsdeskriptoren aus allen Manifestationen eines Werks als Linked Data in einem einzigen Werk-Cluster akkumuliert sind.¹⁴ Auf diese Weise werden sowohl die Auffindbarkeit der entsprechenden bibliografischen Daten erhöht als auch deren Relevanz und damit der Rang bei deren Anzeige beeinflusst. Da jedoch in den WorldCat-Works auch vielfach redundante Daten, insbesondere Schlagwörter, geclustert werden, ist zumindest zum gegenwärtigen Zeitpunkt davon auszugehen, dass die Relevanzmessung partiell dem Zufall der differierenden Beschaffenheit – in diesem Fall der Art und der Breite der sachlichen Erschließung – der gesammelten Daten ausgeliefert bleibt. Das mag zwar nicht überraschen und dürfte sogar eine zwingende Logik implizieren. Es hat jedoch weitreichende Konsequenzen für die dem Art-Discovery-Projekt zugrunde liegende Konzeptidee, eine möglichst verlässliche Fachbibliografie virtuell auf der Basis von Bibliothekskatalogen zu generieren, und zwar ohne Zuhilfenahme traditioneller Parameter wie struktureller Uniformität

und intellektueller Prä-Selektion. Stattdessen verlagern wir in diesem Projekt die Selektions- und Validierungsleistung zu einem nicht unerheblichen Teil auf das Zusammenspiel unserer zur Verfügung gestellten Katalogdaten und auf die Entwicklung von Algorithmen, ohne uns besondere Gedanken darüber zu machen, ob unsere Daten für diesen Zweck überhaupt geeignet sind. Das ist zwar nicht verwunderlich, da bei der Erfüllung der Katalogisierung auf lokaler Ebene die spätere und so entscheidende Interaktion mit anderen Daten bislang nicht präsent war. Es ist aber zumindest bedauerlich, da es den angestrebten Effekt schmälert. In diesem Zusammenhang ist sicher eine Erfahrung von Bedeutung, die im Zuge des Uploads von bislang nicht im WorldCat repräsentierten Bibliotheksbeständen gemacht wurde. Nicht alle Sacherschließungsangaben, die in MARC-Feldern in den WorldCat transportiert werden, kommen auch zur Ansicht und werden für die Indexierung relevant. Es kommt darauf an, wie sie codiert sind und welchem Regelwerk sie entstammen. Nichtnormierte, lokal gepflegte Sacherschließungsangaben, die bei Spezialbibliotheken eine wichtige Rolle spielen können, werden keineswegs automatisch in den WorldCat geladen bzw. in Master-Records integriert, sondern müssen vorab einen für das System des WorldCat erkennbaren ‚source code‘ besitzen.¹⁵ Der WorldCat geht beim Datenupload also durchaus selektiv vor, und man wird ihn sicher zutreffender mit einer komplexen Datenverarbeitungsmaschinerie gleichsetzen können als mit einem gigantischen Aggregator von heterogenen Bibliotheksdaten. Für das Ziel, auf seiner Grundlage so etwas wie eine virtuelle Fachbibliografie aufzubauen, ist die Kenntnis dieser Spielregeln von eminenter Wichtigkeit. Nur so kann man konstruktiv ein Resultat erzielen, das den eigenen Erwartungen entgegenkommt. Es kommt sehr auf die Struktur und Qualität der gelieferten Daten an.

Inhaltsdeskriptoren – Wichtiger denn je

Welche Bedeutung die Datenqualität für die aktuellen und auch zukünftigen Recherchestrategien und Resultatgenerierungen hat und haben wird, ist häufig hervorgehoben worden, nicht zuletzt im Kontext der Rollenzuweisung von bibliothekarisch kontrollierten Katalogdaten in der Linked-Data-Ontologie.¹⁶ Man könnte freilich einwenden, dass die Kunst- und Museumsbibliotheken, die sich zur Zusammenarbeit beim Aufbau des Art Discovery Group Catalogues entschlossen haben, nicht primär von den Linked-Data-Visionen geleitet waren und im Grunde genommen ein recht traditionelles Konzept ei-



nes Verbundkatalogs, wenn auch von moderner Suchmaschinentechnologie ‚gepowert‘, realisieren – zumindest auf dem jetzigen Stand der Projektentwicklung. Auf der anderen Seite hat die prominente Verwendung des Begriffs ‚Discovery‘ einen programmatischen Untergrund, und die Integration eigener und fremder Daten in Kombination mit dem Single-Search-Standard, der darüber hinaus die Ein-Feld-Suche über alle Daten hinweg als Default-Einstellung vorgibt, muss zwingend die Qualität der Daten hinterfragen, wenn auf diese Weise verlässliche und erschöpfende Ergebnisse erzielt werden sollen.¹⁷ Es liegt auf der Hand, dass vor allem die Inhaltsdeskriptoren hierbei eine herausragende Rolle spielen müssen, sollten sie doch im Idealfall jene Informationen enthalten, die den übrigen Daten einer Titelaufnahme eben nicht zu entnehmen sind. Doch die Erfahrung lehrt, dass dies keineswegs immer der Realität entspricht. Es werden nicht nur inhaltliche Angaben, die sich bereits dem Titel des Werkes entnehmen lassen, wiederholt, sondern man findet wenig präzise oder gar keine Inhaltsbeschreibungen. So enthalten einer Untersuchung von 2010 zufolge lediglich 46 % der WorldCat-Titelsätze Schlagwörter und nur 15 % DDC-Klassen.¹⁸ Ganz zu schweigen von den maschinell generierten Aufsatzkatalogisaten aus ArticleFirst, die für das Art-Discovery-Projekt von herausragender Bedeutung sind. Es existiert demnach im WorldCat und damit auch im Gruppenkatalog der Kunst- und Museumsbibliotheken eine beträchtliche Inkonsistenz im Datenbestand, die zu unmittelbaren Konsequenzen bei der Qualität der Ergebnisse beim Retrieval führen muss. Die Single-Search-Methode in Kombination mit der All-Fields-Sucheingabe suggeriert zwar, dass in einem einzigen Schritt alle relevanten Daten erkannt werden, doch ist dies bei themenbezogenen Recherchen kaum möglich, wenn die Inhaltsdeskriptoren die genannte Uneinheitlichkeit und Unzuverlässigkeit aufweisen. Das zuvor erwähnte Relevanz-Ranking wiederum suggeriert, dass die Ergebnisse der Suche entsprechend den Interessen des Nutzers sortiert sind. Auch dies kann angesichts der inkonsistenten Inhaltsbeschreibung nicht geleistet werden, und das Ergebnis wird zusätzlich aufgrund der statistischen Auswertungsverfahren noch stärker verfälscht. Und schließlich suggerieren die Facetten, dass die langen Ergebnislisten sinnvoll und zuverlässig gefiltert werden können, was etwa bei den Publikationsformaten zu befriedigenden Resultaten führen wird, nicht jedoch bei inhaltlichen Kategorien, für die in den Datensätzen keine hinreichend distinkten Identifikatoren existieren. Aber gerade auf die kommt es an, wenn man den Art Discovery

Group Catalogue in erster Linie als ein bibliografisches Instrument versteht, das die Suche nach Informationen und Literatur unterstützen soll, die es erst zu entdecken gilt. Wir müssen uns also in dieser ersten Phase der neuen Partnerschaft mit OCLC eingestehen, dass gegenwärtig noch eine Kluft zwischen den Potenzialen des WorldCat und den von uns gelieferten – oder vorgefundenen – Daten existiert bzw. dass das Aggregieren und Mergen von bibliografischen Daten mit der resultierenden Heterogenität in Qualität und Struktur den Wesensmerkmalen eines Discoverysystems nicht in optimaler Weise entgegenkommt.

Sind nun die beschriebenen Unzulänglichkeiten des Zusammenspiels zwischen der Systemarchitektur einerseits und den geladenen Daten andererseits eher ein bibliothekarisches Problem, mit dem die Nutzer schon umzugehen wissen? Im ungünstigsten Fall wären sie sich dieser Problematik gar nicht bewusst oder – noch schlimmer – es wäre ihnen schlicht egal. Wir haben allerdings wenig Anlass, von einem solchen Nutzerbild auszugehen. Bereits 2009, also in der Frühphase des Aufkommens von Discovery-Systemen im Bibliothekswesen, publizierte OCLC die Ergebnisse einer Untersuchung über die Erwartungen, Forderungen und Erfahrungen von Nutzern verschiedener akademischer Ausbildungsstufen im Umgang mit der WorldCat.org-Datenbank.¹⁹ Unmissverständlich drückt sich in den Ergebnissen die Erwartungshaltung der Nutzer aus, dass die Rankingbildung bei der Anzeige der Treffer Verlässlichkeit und Differenziertheit aufweist, da sie nunmehr, analog zu generischen Suchmaschinen und deren sehr elaborierten Rankingalgorithmen, auch in Bibliothekskatalogen als gültiger Standard betrachtet wird.²⁰ Ebenso werden Inhaltsdeskriptoren – ob nun in normierter oder nicht normierter Form – als ein wichtiges und sachdienliches Mittel der Selektion von Ergebnissen gewertet, sowohl im Hinblick auf Relevanzmessung als auch auf die optionale Kategorienbildung auf der Grundlage von Facetten.²¹ Doch auch bei Filterung der Treffermengen mittels Facetten können verlässliche und konsistente Resultate nur im Falle einer adäquaten Datenqualität erzielt werden, für die die Katalogisate in der aktuellen Gestalt nur zum Teil die geeigneten Voraussetzungen mitbringen. Die zitierte Befragung von 2009 war nicht nur an ‚end user‘ des WorldCat adressiert, sondern auch an Bibliothekare, denen die beschriebenen Defizite verständlicherweise weitaus mehr bewusst sind. Doch nicht allein die Tatsache, dass gut strukturierte Daten auch entsprechend gute Selektierungsmöglichkeiten eröffnen, wird in den Ergebnissen der Umfrage

seitens der Bibliothekare betont, sondern auch die Gefahr benannt, dass Unkenntnis des Qualitätsdefizits der Daten den Nutzern eine irreführende Vorstellung von der Präzision und Zuverlässigkeit der Suchergebnisse gibt.²² Das ist vor allem angesichts der Tatsache bedauerlich, dass das traditionelle, klar geregelte und über Sprachgrenzen hinweg ähnlich strukturierte Modell des bibliografischen Katalogeintrags ausgesprochen gut beispielsweise für das Facetten-Prinzip geeignet ist. Facetten sind Zugangspunkte zum Inhalt der bibliografischen Metadaten, und wenn diese Zugangspunkte nach klar definierten Standards mit regelkonformen Deskriptoren korrespondieren, können die angebotenen Facetten zu sehr präzisen Ergebnissen führen.²³ Doch in dem so entscheidenden Sektor der Inhaltsbeschreibung herrscht die besagte Heterogenität und Unausgewogenheit, und auch die erwähnte, nicht einmal 50%ige Abdeckung der WorldCat-Titelaufnahmen mit Sacherschließungsangaben verschärft das Problem beträchtlich. Muss man sich am Ende also fragen, ob das traditionelle Schema der Bibliotheksdaten, so akkurat es auch gepflegt und international vereinheitlicht, aber so heterogen es andererseits bedient wird, für Suchmaschinen überhaupt geeignet ist? Insbesondere dann, wenn an die Schlüssigkeit der Ergebnisse so hohe Erwartungen geknüpft werden? Oder werden sich alle Probleme in einer Linked-Data-Zukunft von selbst lösen? OCLC hat – wie schon erwähnt – die Konversion der WorldCat-Titelaufnahmen zu Linked-Data-Clustern („WorldCat works“) auf experimentellem Terrain bereits vollzogen und die Informationen aus der MARC-Struktur in eine neue Ontologie der verlinkten Daten auf die Kategorienarchitektur von schema.org gestellt. Freilich ist das Linked-Data-Modell, gerade im Hinblick auf Bibliothekskataloge, keineswegs als Ersatz für Letztere zu verstehen. Vielmehr werden dem Web Informationen ‚mitgeteilt‘, die die Daten ohnehin implizieren, die aber nun von anderen Datenumgebungen außerhalb des Quellenkontextes (also der Katalogumgebung, in der sie entstanden sind und gepflegt werden) verstanden und verarbeitet werden können, womit das Web potenzielle Wissenskorelationen archiviert, die nun in komplexen Suchprozessen erkundbar („Discovery“) werden können. Die Bibliothekskataloge bleiben in diesem System also nicht nur erhalten, sondern sie spielen aufgrund ihres relativ konstanten Informationsinventars im Linked-Data-Paradigma eine prominente Rolle. Und ihre Qualität ist genauso relevant wie in traditionellen Stand-alone-Katalogdatenbanken oder bibliothekarischen Discovery-Umgebungen.²⁴ Was können wir also tun?

Konsequenzen für die Zukunft

Es ist sicher auszuschließen, dass es in absehbarer Zeit zu einer Harmonisierung der Katalogisierungskonventionen und zu international kooperativen Erschließungsprozessen kommt, auch wenn dies in ökonomischer Hinsicht angezeigt wäre und der Effizienz der bibliografischen Discovery-Erfahrung sehr entgegenkäme. Eine solche absichtsvolle Übereinkunft ist vollkommen utopisch. Und auch wenn mutmaßlich immer mehr Bibliotheken weltweit den WorldCat als Katalogisierungsbasis wählen werden, ist die Erwartung einer weitgehend uniformen Katalogisierungspraxis eher unrealistisch. Alle unsere Erfahrungen lehren uns das Gegenteil. Darüber hinaus ist das Einfordern von einheitlichen Standards und Regeln problematisch angesichts der immer größeren, durchaus erwünschten Anreicherung der Bibliothekskataloge und noch mehr der Discovery-Tools mit Metadaten von komplementären Ressourcen, die in formaler und konzeptioneller Hinsicht proprietäre Schemata anwenden können. Auf der anderen Seite wären die Rechercheresultate präziser, verlässlicher in Bezug auf Relevanz und einfacher zu verarbeiten, wenn sie effizienter von entsprechenden Daten unterstützt würden. In dieser Hinsicht steht das Art-Discovery-Projekt ganz am Anfang. Die Metadaten der Partnerbibliotheken, die bislang in diesem neuen Rahmen zusammengeführt werden, interagieren noch nicht perfekt, da sie in lokalen Kontexten entstanden und auf deren spezifische Anforderungen ausgerichtet sind. Diese Autonomie konnten sie, sofern betroffen, unangetastet im virtuellen Katalog artlibraries.net beibehalten, obgleich die Effizienz dieses verteilten Suchmechanismus an der Heterogenität der Daten von Anfang an zu leiden hatte. Im Art Discovery Group Catalogue indes machen alle Beteiligten die neuartige Erfahrung, dass – vereinfacht gesagt – der Nutzen der beigesteuerten Daten entweder hoch oder niedrig sein kann und im ungünstigsten Fall gar kein Mehrwert erzielt wird. Doch das widerspricht dem ursprünglichen Konzept der Initiative, ein innovatives und leistungsstarkes Instrument für die Ermittlung von disziplinbezogenen, vorrangig bibliografischen Informationen zu entwickeln, das im Wesentlichen auf dem Zusammenspiel der Katalogdaten der beteiligten Partnerbibliotheken aufbaut. Dementsprechend kann das Aggregieren der Daten ohne jegliche Modifikation diesem Anspruch kaum genügen. Vielleicht würde schon ein höheres Bewusstsein über die interagierende Rolle der eigenen Daten im WorldCat und die Bereitschaft, Zukunftsvisionen auch jenseits lokaler Selbstgenügsamkeit zu entwickeln, etwas verändern. Wir sind also keineswegs, wie eingangs schon angedeutet, am

Ziel der Initiative angekommen, sondern stehen vor den größten Herausforderungen im Hinblick auf unsere Rolle als Datenproduzenten und Mitspieler beim großen WorldCat-Abenteuer. Gewiss liegt nicht alles in unserer Hand, und manigfache Abhängigkeiten von lokalen, regionalen, institutionellen und funktionalen Anforderungen an unsere Kataloge können die Gestaltungsfreiheit empfindlich einschränken. Auf der anderen Seite wird man das Abweichen von Vorgaben und das Aufgeben von Traditionen immer dann rechtfertigen können, wenn auf diese Weise bessere Resultate erzielt werden. Das Art-Discovery-Projekt eröffnet eine Perspektive, die uns in dieser Hinsicht voranbringen könnte.

1. Der vorliegende Text basiert auf einer übersetzten und überarbeiteten Version meines Vortrags „The Art Discovery Group Catalogue – final goal or starting point?“, gehalten anlässlich des ‚6th artlibraries.net general meeting‘, Kopenhagen, 10.10.2014.
2. <http://artlibraries.worldcat.org/>.
3. http://www.oclc.org/news/releases/2014/201415_dublin.en.html [letzter Zugriff: 10.03.2015].
4. Siehe *The Future of Art Bibliography*. In: *Art Libraries Journal* 36 (2011), 3, Special Issue. Ferner: *Papers from the 5th artlibraries.net general meeting*, Paris, September 27–29, 2012, und 41st ARLIS/NA Annual Conference (Pasadena, CA, April 25–29, 2013), Session: „Building Collective Solutions: the Future of Art Bibliography Initiative (FAB)“, in Auswahl publiziert auf: http://artlibraries.net/archiv_dokumente/archiv_dokumente_de.php [letzter Zugriff: 10.03.2015].
5. Eine informative Zusammenfassung bei: Goncerut Estèbe, Véronique, *De artlibraries.net à Art Discovery Group Catalogue ou l'évolution de métacatalogues pour l'histoire de l'art et d'un réseau des bibliothèques d'art dans le monde* (http://www.ressi.ch/num15/article_103 [letzter Zugriff: 10.03.2015]).
6. Programm zum EMEA Regional Council Meeting, Florenz, 10./11. Februar 2015, S. 15.
7. EMEA = Europe, Middle East, Africa (<http://www.oclc.org/membership/councils.en.html#1480337839> [letzter Zugriff: 10.03.2015]).
8. APAC = Asia-Pacific
9. Siehe hierzu: <http://www.oclc.org/support/help/batchload/Default.htm> [letzter Zugriff: 10.03.2015].
10. Betroffen sind insbesondere nicht normierte, in Spezialbibliotheken jedoch durchaus vorkommende Sacherschließungsangaben. Vgl. hierzu Rüdiger Hoyer, „kubikat in a transatlantic context“, Vortrag, Kopenhagen, 10.10.2014, http://artlibraries.net/archiv_dokumente/dokumente/Copenhagen_Pres2_text_hoyer.pdf [letzter Zugriff: 10.03.2015].
11. OCLC macht es einem nicht gerade leicht, die Prinzipien der Datenverarbeitung im WorldCat zu begreifen. Etliches erschließt sich nicht von selbst. Auch die zugehörige Dokumentation ist nicht leicht auffindbar. Die Website von OCLC ist hierbei nur von begrenzter Hilfe.
12. Vgl. <http://www.oclc.org/support/services/worldcat-local/faq/search.en.html> [letzter Zugriff: 10.03.2015].
13. Das OCLC-Projekt GLIMIR (Global Library Manifestation Identifier), begonnen 2009, zielt in diesem Kontext auf zwei Effekte: einerseits auf das Clustern von Katalogdaten verschiedener Manifestationen eines und desselben Werkes und andererseits auf die Zusammenführung von Metadaten aus verschiedenen Quellen, die sich aber auf dasselbe Werk beziehen, um die Effizienz des Discovery-Prozesses zu erhöhen. Vgl. Gatenby, Janifer u. a., *GLIMIR: Manifestation and Content Clustering within WorldCat*. In: *Code4lib journal* (2012), 17 (<http://journal.code4lib.org/articles/6812> [letzter Zugriff: 10.03.2015]) und Thornburg, Gail, *A Candid Look at Collected Works. Challenges of Clustering Aggregates in GLIMIR and FRBR*. In: *Information Technology & Libraries*, 33 (2014), 3, S. 53–64.
14. Vgl. <http://www.oclc.org/data.en.html> [letzter Zugriff: 10.03.2015].
15. Die Zuteilung eines source code obliegt exklusiv der Library of Congress. Dieser Code wird in den 650er-MARC-Feldern im Unterfeld \$2 hinterlegt. Nur wenn die MARC-Felder mit dem zweiten Indikator 7 eingeleitet werden, wird die enthaltene inhaltbeschreibende Information für den WorldCat verarbeitet. Vgl. <http://www.loc.gov/marc/bibliographic/bd650.html> und <http://www.loc.gov/standards/source-list/index.html> [letzter Zugriff jeweils: 10.03.2015].
16. Zuletzt etwa bei Gonzales, Brigid M., *Linking Libraries to the Web. Linked Data and the Future of the Bibliographic Record*. In: *Information Technology & Libraries*, 33 (2014), 4, S. 10–22, bes. S. 14.
17. Vgl. Prescott, Leah, *Single Search. The Quest for the Holy Grail*, OCLC Research 2011, siehe <http://www.oclc.org/research/publications/library/2011/2011-17.pdf> [letzter Zugriff: 10.03.2015], S. 18.
18. Smith-Yoshimura, Karen u. a., *Implications of MARC Tag Usage on Library Metadata Practices*, 2010, S.19–20, siehe www.oclc.org/research/publications/library/2010/2010-06.pdf [letzter Zugriff: 10.03.2015].
19. Calhoun, Karen u. a., *Online Catalogs. What Users and Librarians Want. An OCLC Report*, 2009, siehe www.oclc.org/reports/onlinecatalogs/fullreport.pdf [letzter Zugriff: 10.03.2015].
20. Ebd., S. 11, 14 und 54. Siehe auch Kinstler, Till, *Making Search Work for the Library User*. In: *Catalogue 2.0. The Future of the Library Catalogue*, hg. von Sally Chambers, London 2013, S. 25–26.
21. Calhoun (wie Anm. 19), S. 16, 18.
22. Ebd., S. 40.
23. Vgl. Kinstler (wie Anm. 20), S. 30.
24. Vgl. den realistisch-bescheidenen Begleittext zum Projekt „Linked Data for Libraries (LD4L)“ von Dean B. Krafft („Why Linked Data?“): <https://wiki.duraspace.org/pages/viewpage.action?pageId=43910411> [letzter Zugriff: 10.03.2015].