

# BIG DIVA



*Redefining (Re)Search*

*A SERVICE OF THE ADVANCED RESEARCH CONSORTIUM*



**arc**

*HOSTED BY TEXAS A&M UNIVERSITY*





## BIGDIVA

BigDIVA (Big Data Infrastructure Visualization Application) is a web-based search and discovery service designed for humanities scholars and students. A complete research environment for humanists, BigDIVA gathers together the highest quality open-access digital resources alongside proprietary materials (described in the next section, “ARC Catalog,” p. 3). The BigDIVA search interface can be accessed via a web browser for use on large screen installations as well as by individual computers and laptops. By subscribing to BigDIVA, libraries are investing in a tool that legitimizes open-access publishing and crosses proprietary and open-access boundaries, situating open educational learning resources and scholarship within a search tool that accesses the most relevant resources of all kinds. Beyond this, libraries will also be supporting the Advanced Research Consortium (ARC), an organization that fosters the creation and dissemination of peer-reviewed digital research projects (please see “The Advanced Research Consortium,” p. 6).

## ARC CATALOG

The BigDIVA web service ([bigdiva.org](http://bigdiva.org)) is fed by the ARC Catalog, a resource developed by scholars for scholars. As mentioned above, the catalog includes open-access resources: peer-reviewed digital projects produced and evaluated by field-specific specialists. It combines these with proprietary materials, including journals and books from JSTOR and ProjectMUSE, as well as rare book catalog records from the best collections in a given field, including the English Short Title Catalog and the North American Imprints Program. The catalog also includes metadata and full text from Gale-Cengage Learning’s Eighteenth-Century Collections Online, and metadata and full text from ProQuest’s Early English Books Online. Full text has been provided by the Text Creation Partnership, and more is coming from the Mellon-funded Early Modern OCR Project ([emop.tamu.edu](http://emop.tamu.edu)). Proprietary and open-access educational materials are fully represented in the ARC catalog.



## THE BIGDIVA SERVICE

ARC brings together a group of faculty from research institutions to solve problems arising from digital database design that impede humanistic research. In BigDIVA your students and researchers can search and visualize the proprietary collections and databases to which your institution subscribes alongside open-access educational resources; digital artifacts, and scholarship. These items are searchable via shared, discipline-specific taxonomies that make significant research possible. Faculty and students can collaborate in research projects using BigDIVA on their phones, laptops, desktops, or large screen installations. Whether working singly or in groups, users can more easily comprehend search results of large amounts of humanities data.

Beyond merely searching across the entire ARC catalog, BigDIVA *visualizes* search results, allowing scholars to see at once the number of results provided by their current search parameters *as well* as the number of results available to them should they make different choices. As one user reported,

*Using BigDIVA to search the ARC Catalog takes the onus off scholars and students to start with a specific research question, instead allowing questions to arise naturally as they explore the catalog. It is not simply a search interface but also a discovery tool.*

While ARC is dedicated to providing BigDIVA as a free service to those wishing to search open-access materials, institutional subscriptions to BigDIVA unlock the ability to seamlessly explore each of the supported, proprietary databases licensed by that institution along with those freely available resources. Our service also provides a help desk that supports all users, in the process improving metadata and refining the search tool based on user input.

BigDIVA provides users with:

- A method for effectively handling large numbers of search results by representing them spatially;
- An alternative to search interfaces that rely on static lists of results;
- A way of revealing to researchers how their search parameters could be adjusted in order to garner better results;
- A way to introduce serendipity into search and research.

Therefore, BigDIVA augments research capacity because it is not limited by preconceived research questions.



## BENEFITS TO SCHOLARSHIP

BigDIVA's interface provides three significant advantages for scholars compared to the traditional search engines that have come to dominate library catalog interfaces.

### **BigDIVA allows users to search high-quality, open-access resources alongside proprietary ones.**

The ARC Catalog, as visualized through the BigDIVA service, includes metadata from leading humanities proprietors. Although online library catalogs allow users to search these resources from a single interface, institutions at times struggle to promote and provide access to a growing collection of freely available, high quality digital resources, especially scholar-generated digital projects and open educational learning resources. ARC peer reviews and aggregates such resources alongside proprietary content, enabling humanities scholars to discover the latest high quality scholarship regardless of its source or purveyor.

### **BigDIVA removes many of the biases that often cloud scholarship and discovery due to search engine ranking algorithms.**

All search engines, including Google and library catalogs, utilize algorithms to sift through thousands and even millions of search returns. While these algorithms may be effective for discovering known entities (questions for which there is a recognized answer), they hinder original research. It is now accepted knowledge that these algorithms introduce bias into search results (whether based on citations, click-throughs on similar search queries, or other metadata), making it difficult for scholars to see outside the constraints of their expectations. The search tool for BigDIVA only categorizes the content, allowing users – not computer algorithms – to determine which items are interesting and relevant.

### **BigDIVA transforms the act of search into research.**

Minimization of bias introduces serendipity into digital research. BigDIVA can display all possible results at once, yet also reveal the items available for *unselected* dates and categories. That is, users will be able to see the returns that lie outside of their search constraints, and thus outside their expectations. BigDIVA elevates the act of search to research.



## THE ARC/BIGDIVA INNOVATION

We are requesting libraries to participate in a new initiative that addresses three questions:

1. Is it possible for scholars to do research, as the ARC Executive Committee is continuously doing — research into the nature of data; research into how digital work transforms disciplinary categories, conceptions, and assumptions; research with data; research into the nature of search design — while transmitting the product of this research directly to libraries? That is, scholars cannot and should not become vendors, who are very good at what they do! Can we successfully disseminate our product while focusing primarily on continuous research and development, ignoring marketing, for example?
2. Is it possible for scholars to help transform the publishing-purchasing economy for scholarly work to render it less market-driven? Our goal is not to compete with and undercut vendors — we need them — but to propose an economy in which libraries can help each other save money, directly lowering the costs of resources for each other by purchasing them (see p. 7, note \*\*).
3. Is it possible for scholars to help transform development of search-and-discovery tools so that they are not similarly market-driven? That is, vendor-developed tools and interfaces aggregate proprietary resources, sometimes their own, sometimes in collaboration with other vendors: can we insert into this development a process whereby scholar-approved and scholar-developed open access materials find entry?

## THE ADVANCED RESEARCH CONSORTIUM (ARC)

ARC ([ar-c.org](http://ar-c.org)) developed out of the NINES project ([nines.org](http://nines.org)) at the University of Virginia. In 2003, scholars led by Jerome McGann and Bethany Nowviskie met to create an organization that would not only aggregate together digital projects produced by nineteenth-century scholars – the Blake Archive, the Rossetti Archive, Whitman – but also provide a means for evaluating digital projects to come. As scholars clamored for similar organizations that spanned more than the nineteenth century, ARC came into existence to support them all (see Appendix 2, pp. 10-11). This consortium was thus born out of a desire to provide peer review for digital work that scholars were doing, a desire to legitimize but also to provide guidelines for successful and meaningful implementations of digital scholarship.



## INITIAL FOUNDING LIBRARIES

The BigDIVA service is being offered to libraries directly, with no commercial organization mediating between the user-producers of the data (faculty and students), on the one hand, and libraries on the other. The funds collected from annual subscriptions to the BigDIVA service go toward sustaining ARC (budget available upon request).

A select group of libraries are invited to become *founding members* of the BigDIVA service to support ARC over the next three years in exchange for multiple benefits. Founding Libraries may:

- Receive free instance(s) of COLLEX for special collections at your library (please see Appendix 3 for more information);
- Guide development of ARC and BigDIVA by submitting enhancement requests;
- Attend semi-annual meetings to participate in decision-making discussions (optional);
- Host one of the semi-annual ARC meetings (in exchange for a reduced fee) during which members of the ARC Steering Committee can offer public talks or panels;
- Schedule workshops about using BigDIVA for research and teaching to faculty and students at your institution, including a workshop for faculty on creating peer-reviewable digital editions and resources;
- Receive free subscription after an initial two (2) year term.

### **Founding Library Contribution**

\$20,000 per year\* for two years  
Free thereafter

(Libraries wishing to continue to participate per above may pay \$10,000 per year after the initial two years, \$5,000 if hosting meetings.)

### **Partner Library Contribution**

\$2,500 for the first year  
\$250 per year thereafter

\* If hosting an ARC meeting, the Founding Library will contribute \$15,000 and is responsible for \$5,000 in meeting expenses during that year.



## INTERESTED?

View *New Modes of Scholarly Research: How Libraries Can Help* ([bit.ly/ARC-Libraries](http://bit.ly/ARC-Libraries))  
a presentation by Dr. Laura Mandell for the Dean's Research Lecture Series at the University  
of Saskatchewan Libraries, Fall 2015.

### **Contact Information:**

Advanced Research Consortium  
c/o Center of Digital Humanities Research  
Attn: Dr. Laura Mandell  
Texas A&M University  
4213 TAMU  
College Station, TX 77843-4213

**Tel:** +1 (979) 845-8345

**Fax:** +1 (979) 862-4334

**Email:** [codhr@tamu.edu](mailto:codhr@tamu.edu)

Laura Mandell, Ph.D.  
Director, Advanced Research Consortium  
Professor, Department of English  
Director, Center of Digital Humanities Research  
Texas A&M University





## APPENDIX 1: TECHNICAL SPECIFICATIONS

Relying on W3C specifications for creating linked data – specifically the Resource Description Framework (RDF) – we have made interoperable and simultaneously searchable humanities databases of primary texts and databases of research publications. The scholar-generated metadata categories established by the Advanced Research Consortium (ARC), which is continuously in the business of evaluating and setting metadata standards, makes this infrastructure possible. Contributors create metadata according to ARC RDF requirements, and then this metadata is indexed in a dedicated Solr server (the same software used by Netflix), which means two things:

1. The lean ARC metadata adheres to library standards such as Dublin Core and Library of Congress authorities, but it adds categories – emerging, nuanced versions of genre and discipline designed to interact together as facets that will return the search results most relevant to humanities researchers.
2. ARC metadata and full-text (where available) are searchable through a Lucene search engine. Apache Solr features unparalleled efficiency in searching large datasets, which means users can search ARC's over 1.7 million digital artifacts in moments.

BigDIVA is a JavaScript program that organizes search returns visually according to ARC's metadata elements. Since JavaScript has become a web interface design standard, BigDIVA can be used with major browsers. BigDIVA visualizes all free-culture resources freely, but it requires an institutional subscription to visualize and search proprietary resources. When an institution purchases the BigDIVA service, the proprietary catalogs to which that institution subscribes are made available via IP Address detection.



## APPENDIX 2: THE ADVANCED RESEARCH CONSORTIUM

The ARC Catalog was generated by a community of scholars that began to approach digital aggregation in 2003. In that year, Jerome McGann organized a steering committee to launch his Mellon-funded project, NINES, or the Networked Infrastructure for Nineteenth-century Electronic Scholarship ([nines.org](http://nines.org)). NINES served as a model for launching 18thConnect.org — focused on the preceding century — and the Mellon-funded MESA community, the Medieval Electronic Scholarly Alliance. This group of field-specific scholarly communities and online-finding aids needed to create an overarching organization that would make available for other groups the technical model spearheaded by NINES, provide infrastructure and programming support, and negotiate on behalf of all the groups with vendors that might contribute data. ARC now sustains the Medieval Electronic Scholarly Alliance ([mesa-medieval.org](http://mesa-medieval.org)), the Renaissance Knowledge Network ([rekn.org](http://rekn.org)), 18thConnect ([18thconnect.org](http://18thconnect.org)), NINES ([nines.org](http://nines.org)), Modernist Networks ([modnets.org](http://modnets.org)), the Canadian Writing Research Collaboratory ([cwrc.ca/en](http://cwrc.ca/en)) and Studies in Radicalism Online ([studiesinradicalism.org/](http://studiesinradicalism.org/)). Other participants are forthcoming, the Networked Early American Resources finding-aid (NEAR), sponsored by the American Antiquarian Society and New York University, and Musical Scholarship Online (MuSO). These are scholarly communities that peer review digital projects in field-specific disciplines (medieval studies, renaissance studies, eighteenth-century studies, nineteenth-century studies, modernist studies, Canadian writings, radicalism studies, early American studies, and music, respectively).

Those online finding aids and scholarly communities contribute metadata to ARC for their peer-reviewed projects. ARC then integrates this metadata with the proprietary resources deemed relevant by each community. The ARC server feeds this data into the online finding aids for each of these communities as well as into the BigDIVA search and discovery service, bringing together more than 1.7 million cultural heritage objects spanning the history of Western civilization from the medieval to the modern period.

ARC metadata for these cultural materials is distinctive, complementing extant library metadata. Our categories are designed to meet the needs of humanities scholars for the purposes of search and discovery: our principles of metadata reform and our current categories are freely available online ([wiki.collex.org](http://wiki.collex.org)). These categories evolve as new communities join ARC and are all thoroughly discussed at ARC meetings that bring scholars and librarians together. Our goals in establishing cataloging taxonomies are to use library standards such as Dublin Core integrated with new linked open data requirements. Both of those goals are subordinated to our primary mission: to keep pace with emerging needs as traditional and avant-garde scholars confront this new data deluge. ARC metadata provides a discovery layer that makes sense to researchers seeking digital materials. It functions like an index of what is available digitally that has been vetted and approved by experts in the fields



*Appendix 2: The Advanced Research Consortium continued...*

of literary, historical, and cultural studies. ARC's metadata categories reflect research interests of our current time as well as the terminology that is now emerging to represent new digital work.

ARC holds two meetings per year: two directors from each scholarly community and often their project managers come together to discuss state-of-the-art digital projects, digital pedagogy, peer-reviewing policies, and especially ARC metadata design.



## APPENDIX 3: COLLEX FOR YOUR SPECIAL COLLECTION

In 2015, Michigan State University Libraries contacted ARC because they wanted to set up an instance of COLLEX, the Collection and Exhibit Tool originally developed at the University of Virginia by Bethany Nowviskie and Eric Hatcher, that was the prototype for BlackLight. Librarians from Michigan State University Libraries Special Collections wanted to be able to present to the public their collection of materials on political radicalism\* but wanted to place it into a complete research environment — that is, to make it searchable along with relevant journal articles and other proprietary resources as well as metadata from the Digital Public Library of America (DPLA). After struggling on their own to do so, they turned to ARC, and we developed “Studies in Radicalism Online” (SiRO; <http://www.studiesinradicalism.org>).

The development of SiRO was fueled by a desire

- to make readily available and findable the items in several collections, beginning with “Sacco and Vanzetti”;
- to provide a virtual community for those engaged in the study of radicalism;
- to meet emerging needs of this community, including making recommendations on technological and scholarly standards for electronic scholarship, the aggregation of data, and the ability to discover and repurpose this data; and
- to aggregate resources (both primary and secondary) relevant to radicalism studies.

Like Michigan State University Libraries, Founding Libraries may request one or more COLLEX installations to design in their own way, a process for ingesting metadata, the capacity to request that ARC aggregate related materials (we have contracts with JSTOR, ProjectMUSE, GaleCengage Learning, ProQuest, etc., procedures in place for ingesting metadata from the DPLA and Europeana, and can contact other proprietors and organizations to create new contracts). ARC’s Executive Board (<https://tinyurl.com/ar-cExecBd>) will release a call among faculty interested in Digital Humanities to direct your community, if desired.

\*The extent of the MSU Collections can be seen by going to <http://www.studiesinradicalism.org/search>, expanding “Other Digital Collections” on the right-hand side of the search browser, and then expanding “Michigan State University Libraries.”



## NOTES



BigDIVA