# Faceted Browsing for Combined Access to a Digital Repository and a Library Catalog

Bess Sadler
Leslie Johnston
University of Virginia Library



DLF Fall 2007 Forum

### What is Project Blacklight?

- Blacklight is a research project. Bess Sadler and Erik Hatcher started an experiment to index MARC data in Lucene/Solr in January 2007.
- The first prototype included ~3.7M MARC records, 320 Tang Dynasty Chinese poems in TEI, and 470 Digital Collections Repository objects.
- MARC indexing consulting came from Erin Stalberg (formerly UVA but now at NC State) and Ed Summers at Library of Congress.
- Why "Blacklight"? Solr? UVA? Blacklight. Get it?
   Credit (or blame) Erik Hatcher.

### What Blacklight Can Solve For Us

- Faceted browsing where we have none.
- Relevancy ranking where we have none.
- Facilitates the use of search and faceted browse together to make it easier to perform complex discovery operations without knowing the logic behind it.
- In addition to providing access to the entire catalog, we can create
  multiple additional interfaces to accommodate specialized browsing of
  different types of collections, such as music collections (audio and
  scores), that takes advantage of specialized use of MARC.
- We can mix in data that's not explicit in the MARC record. In our music interface, we solved a user frustration where they couldn't find music by century – we're extrapolating at the time of indexing from the year that is in the metadata.
- We can federate indexes of MARC records with indexes of metadata for digital objects in our Repository for a single discovery method.
- We can potentially add additional applications into the mix, such as one where we will track our holdings that are included in Google Book Search that will supply link URLS for the interface.

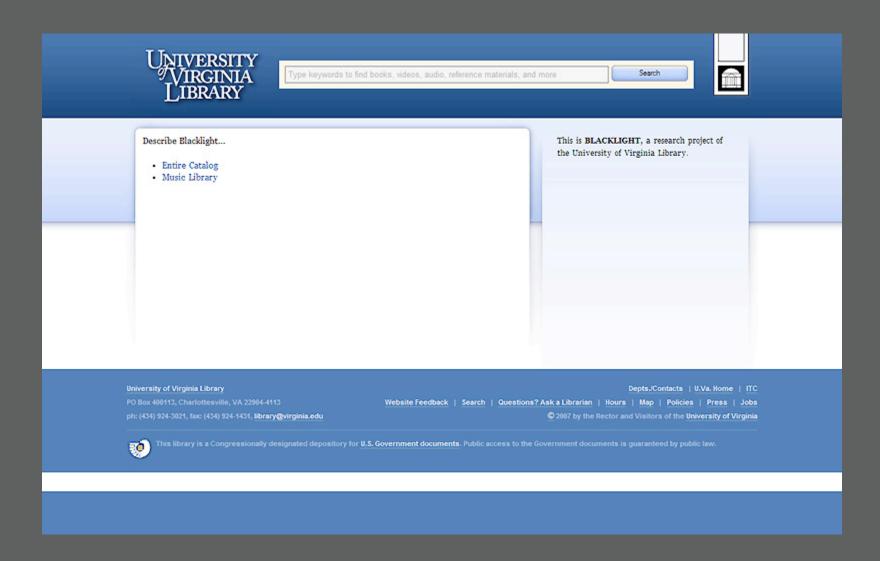
### Blacklight Technical Details

- The MARC records are indexed through a Ruby script, directly from MARC binary without transforming into MARCXML first.
- The implementation was accomplished using Ruby on Rails.
- The box it's all running on is a development server, with four medium fast CPUs and 3.5Gb of RAM.
- Solr Flare is running through Jetty, and Blacklight is running on a pack of four mongrel instances, with apache doing load balancing out front.
- Sessions get written to a MySQL database, and we use Capistrano 2.0 for deployment and versioning.
- It sends us all an email every time an error gets generated.

### Issues in Project Blacklight

- Our MARC data isn't utf-8-compliant and that caused issues with the diacritics.
- Indexing in new ways always exposes new data inconsistencies.
- We have not yet identified a production workflow for keeping the catalog updated daily.

### Blacklight Portal Entry



## Combined Catalog with MARC Records and Repository Objects



Language: English

Year: 1920

Date: [c1920-1922].

Creator: Bruce, Philip Alexander; Macmillan

shadow of one man,

Type keywords to find books, videos, audio, reference materials, and more

Search



o filter(s) applied. 288165 items total.

Page 1/28817

← 1-10 of 288165 →

History of the university of virginia, 1819-1919;the lengthened shadow of one man,

Collection: UVa Library Text

Subject genre: Jefferson, Thomas, 1743-1826; University of Virginia; History

Composition era: 20th Century

Content type: TEI

Source: Digital Library

Title: History of the University of Virginia, 1819-1919;; the lengthened

This is <u>BLACKLIGHT</u>, a research project of the University of Virginia Library.

### SOURCE

Digital Library (10143) Library Catalog (153692) Music (13030)

Other (111300)

### FORMAT

Archives (1683) Audio CD (1091) Book (215690) CD-ROM (307) Cassette (304)

DVD (3) Digital Media (8)

### Music Facets for Browsing



Type keywords to find books, videos, audio, reference materials, and more



Search

o filter(s) applied. 522578 items total. Page 1/52258 - 1-10 of 522578 -On trial; the soviet state versus "abram tertz" and "nikolai arzhak." Ter□t□s, Abram, Alderman: LAW Language: English Bibliography of ghana, 1957-1960 Library: Alderman Subject geographic: Language: The composition and origin of the terrestrial fauna of middle asia Author: Kryzhanovski□i, O. L. Library: Brown SEL Topic form genre: Zoology Subject geographic: Asia, Central Language: English A general catalogue of books, arranged in classes, offered for sale Quaritch, Bernard, Library: Alderman Catalogs, Booksellers'; Books; Rare books; Prices Topic form genre: Great Britain Subject geographic: The new universal gazetteer and geographical dictionary Thomson, John

This is BLACKLIGHT, a research project of the University of Virginia Library. THIS TEXT IS COMING FROM THE MUSIC VIEW HELPER RECORDINGS AND SCORES Recordings (13599) Recordings or Scores (18942) Scores (5343) COMPOSITION ERA 10th Century (2) 11th Century (646) 12th Century (20) 13th Century (14) 14th Century (24) 15th Century (49) 16th Century (876) 17th Century (332) 18th Century (1351) 19th Century (1477) 1st Century (9) 20th Century (2568) 21st Century (63) 5th Century (3) 6th Century (103) 7th Century (2) 8th Century (3) 9th Century (3) LIBRARY Alderman (276447) Astronomy (1050)

### More Music Facets for Browsing

Library: Topic form genre: Subject geographic: Language: lvy Annex Libraries California, Southern

English

### Memorials of a quiet life

Author: Hare, Augustus J. C Library: Alderman Language: English

### Treaties and alliances of the world

Library: Alderman; LAW
Topic form genre: Alliances; Treaties
Language: English

o filter(s) applied. 522578 items total. Page 1/52258 ← 1-10 of 522578 →

### RECORDING FORMAT

CD (2837) Cassette (496) DVD (4) LP (3728) Open Reel Tape (462) VHS (2476) Video Disc (106)

### LANGUAGE

(18773)d] (1) 536 (1) ??? (822) Afrikaans (119) Afroasiatic (Other) (3) Akkadian (3) Albanian (19) Altaic (Other) (1) Amharic (1) Arabic (2030) Aramaic (4) Armenian (8) Assamese (12) Austronesian (Other) (6) Avestan (2) Awadhi (4) Aymara (1) Azerbaijani (1) Balinese (1)

### INSTRUMENT

Alto voice (54)
Band (27)
Baritone horn (3)
Baritone voice (45)
Bass clarinet (16)
Bass voice (45)
Bassoon (109)
Bowed Strings, Other (2)
Bowed Strings, Unknown (1)
Bowed Strings, Unspecified (4)

### And Even More

### TOPIC FORM GENRE

History (28137)

Electronic journals (26681)

History and criticism (13026)

Law and legislation (7092)

Education (4640)

Research (3797)

Finance (3591)

Philosophy (3536)

Study and teaching (3254)

Architecture (3171)

Women (3061)

World War, 1939-1945 (2853)

Music (2784)

Social aspects (2662)

Environmental aspects (2455)

Economic aspects (2335)

Insurance, Flood (2327)

Management (2262)

Government policy (2195)

English language (1957)

### SUBJECT GEOGRAPHIC

United States (70786)

History (28614)

Politics and government (16587)

Great Britain (9938)

France (7580)

Italy (6735)

India (6690)

Germany (6341)

Virginia (6334)

Economic conditions (5963)

Foreign relations (5613)

Soviet Union (5309)

Description and travel (5053)

China (4158)

Europe (4154)

Social conditions (4154)

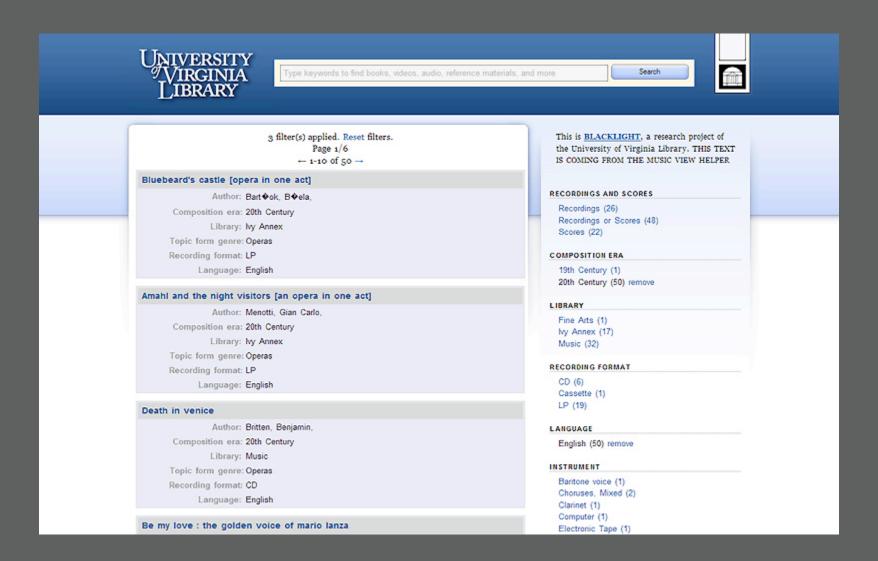
England (4102)

Social life and customs (3394)

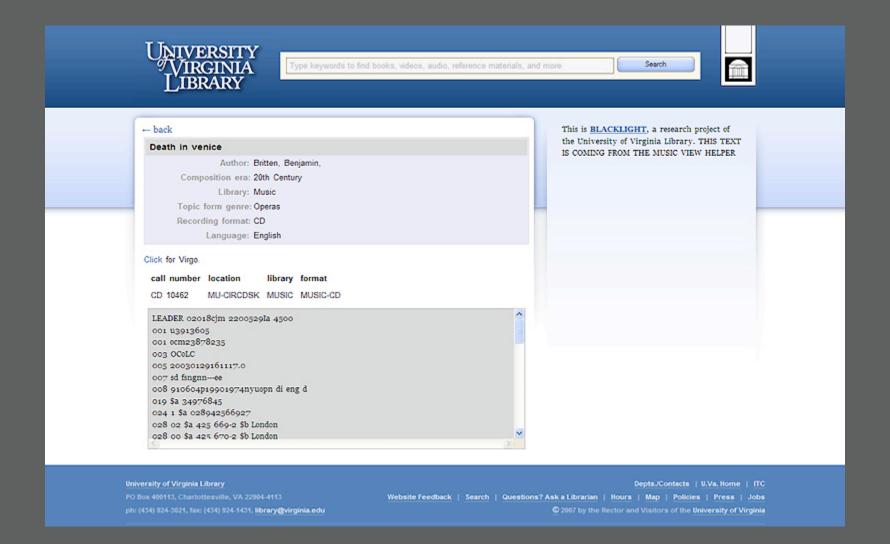
Spain (3147)

Civilization (3073)

### 20th-century Opera in English



### View a Record



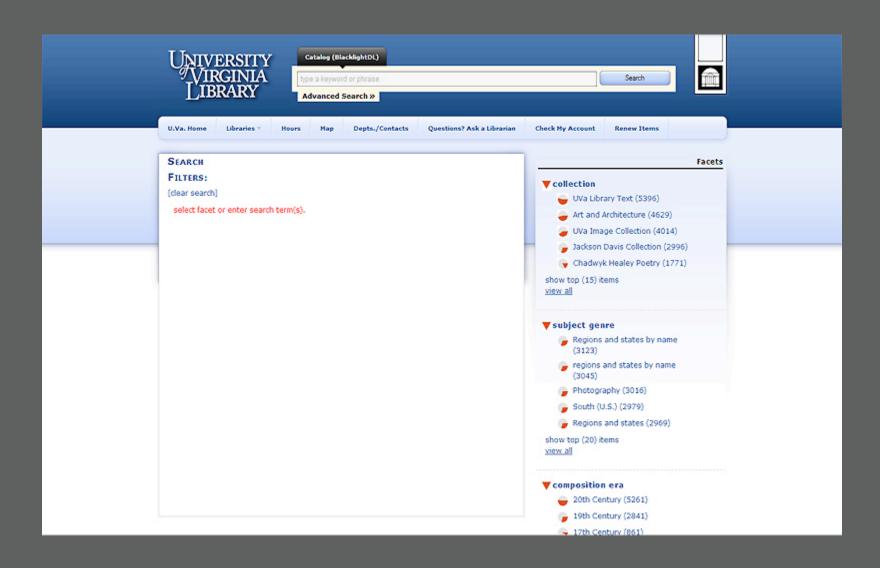
## Books on Africa for Semester at Sea: Alternate Interface for a Shadowed Record Set



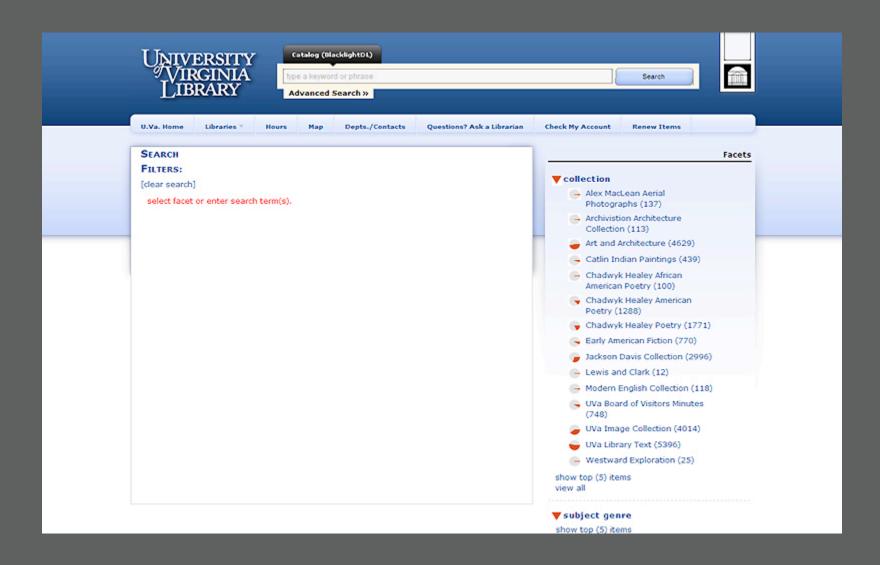
### BlacklightDL

- A corollary project to replace the XPAT indexes and search interface for our Digital Collections Repository (on top of Fedora).
- Indexes full text and metadata for over 10,000 TEI texts, almost 4,000 EAD, and over 20,000 images.
- Work done by OpenSource Connections in conjunction with Bess Sadler and Matt Mitchell at the UVA Library.
- BlacklightDL will be folded into Blacklight.

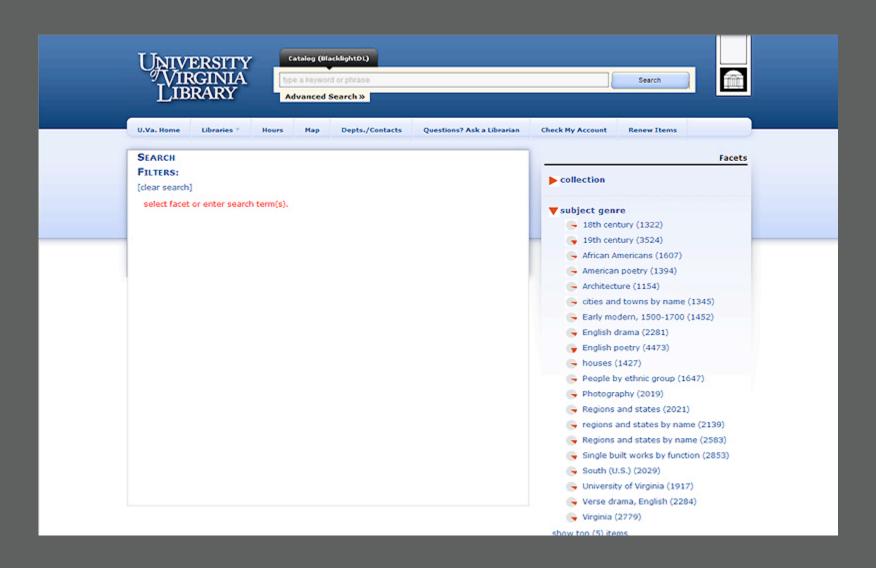
### BlacklightDL Main Screen



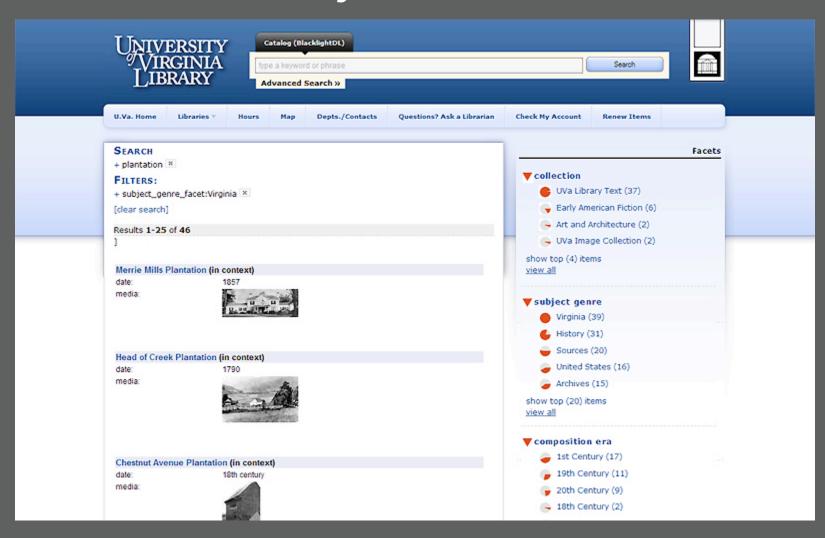
### BlacklightDL Collection Browse



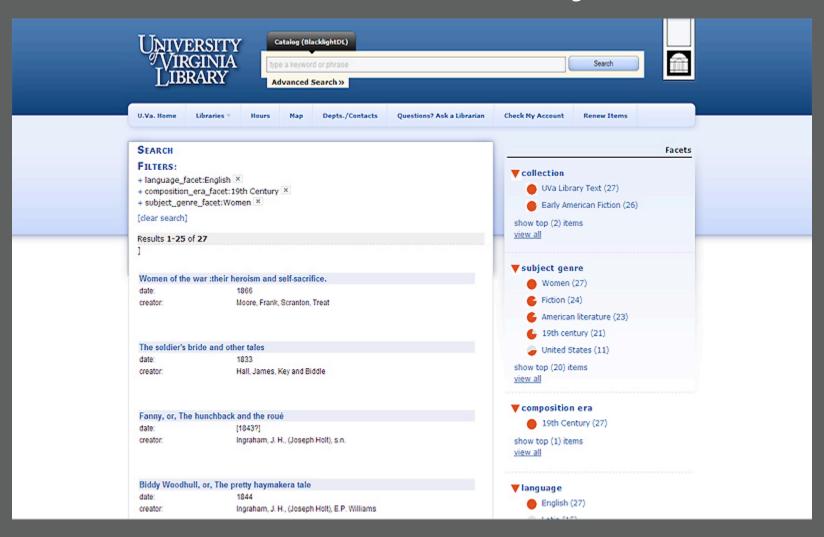
### BlacklightDL Subject Facet



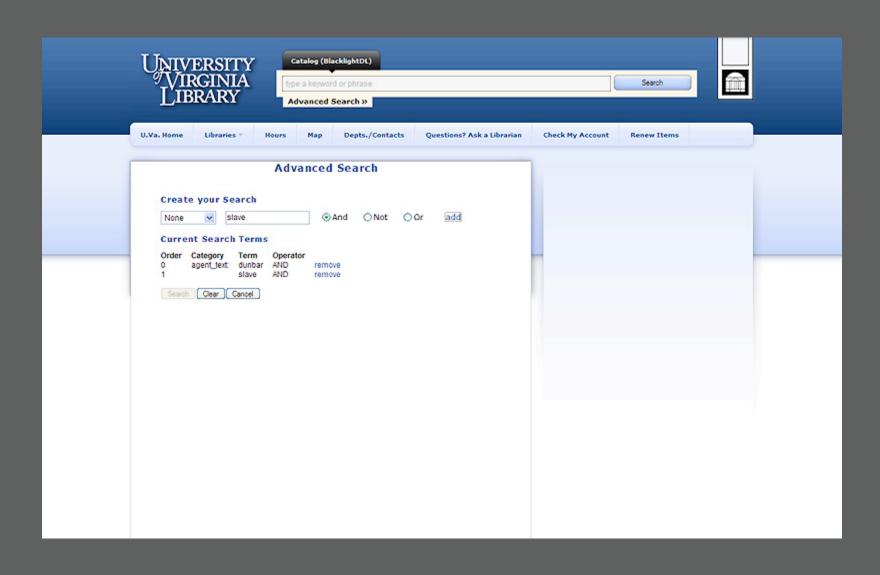
## BlacklightDL Search Plus Subject Facet Filter



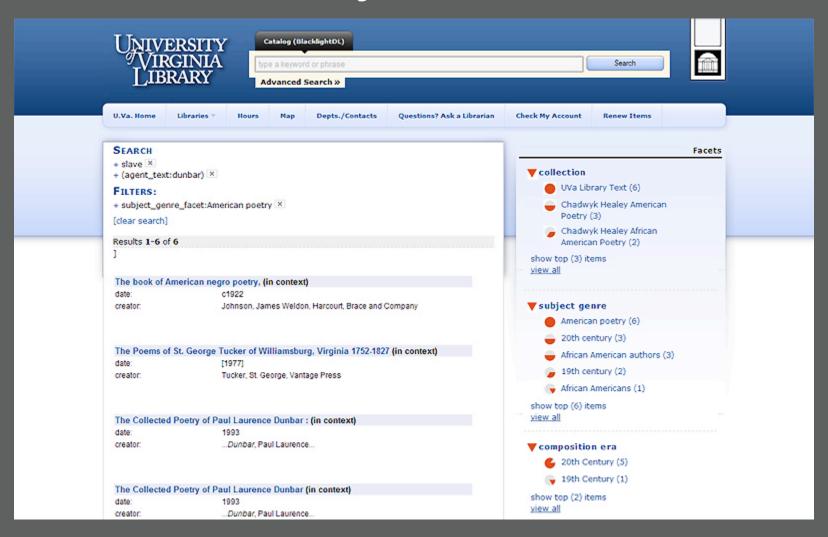
## 19<sup>th</sup>-century Books in English with Women as a Subject



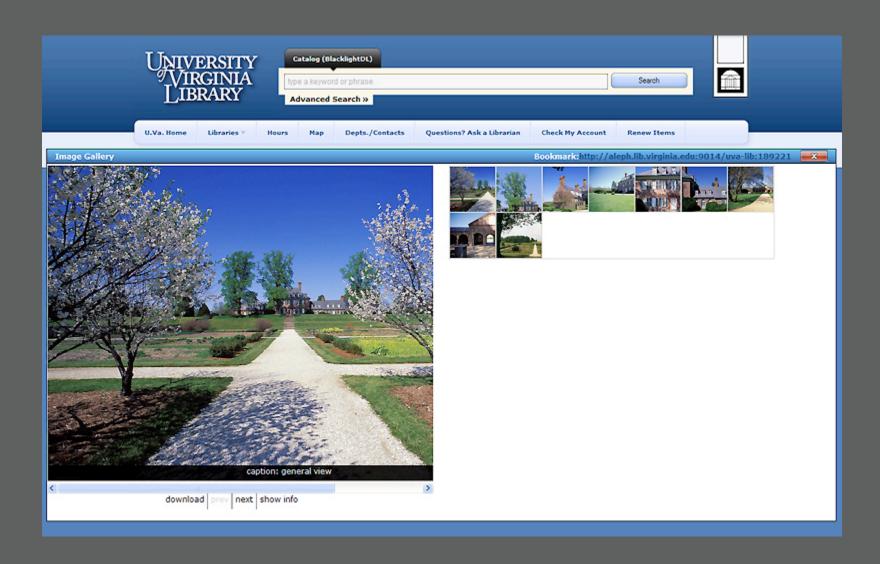
### Advanced Search in BlacklightDL



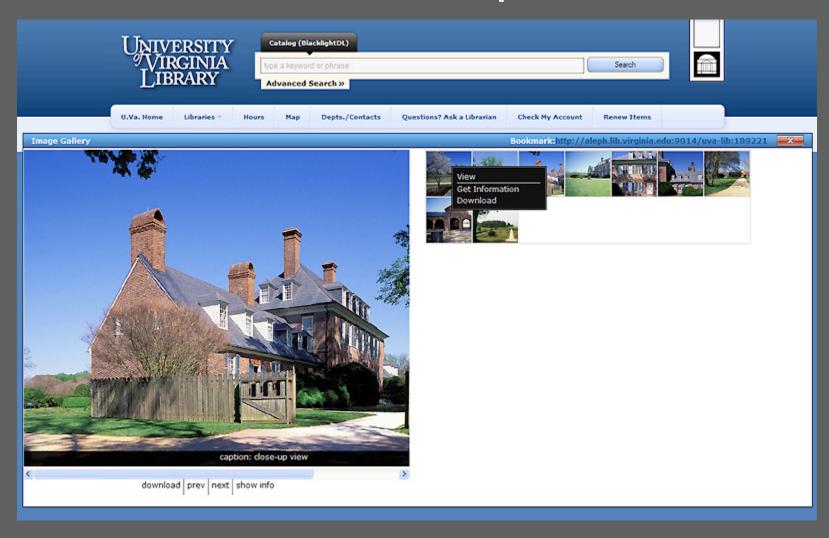
## The Advanced Search Results Plus a Subject Facet Filter



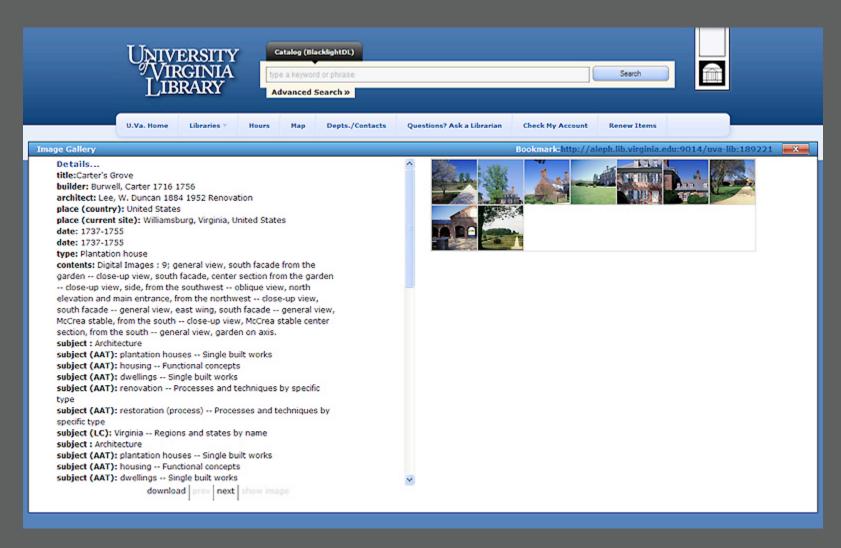
### BlacklightDL Image Object View



## BlacklightDL Image Object View with Menu Options



## BlacklightDL Image Object View with Metadata



### Questions?



Contact us:

bess@virginia.edu johnston@virginia.edu