5-2016

# Safeguarding Your Content with Digital Commons

bepress

# Safeguarding Your Content with Digital Commons

Version: May 2016
Available at http://digitalcommons.bepress.com/reference/18

## A Comprehensive Approach

Over time, many different technologies may be used to safeguard an institution's intellectual assets. As an organization responsible for providing hosted repository services, we see our role as integral to helping you achieve your long-term goals in this area.

Our current strategy for safeguarding your content entails a robust infrastructure, archival options, regular maintenance schedule, and an array of services that we know many individual institutions find difficult to provide for themselves. We protect your intellectual assets against a variety of threats, from the passage of time to direct, malicious attacks to the service. We take these threats seriously and constantly evaluate the level of security our system provides.

We also recognize the importance of our ability to maintain a high level of service to our subscribers. Since our founding in 1999, bepress has been very successful at providing software services to the scholarly community. We have been developing and actively hosting institutional repositories since 2002. During this time, we have invested considerable resources in developing Digital Commons into the world's leading hosted repository platform. Today we host repositories for over 400+ institutions around the world. We are financially healthy, and our revenues are steady.

Digital Commons is a high-availability platform that utilizes a distributed file system to provide fast and reliable access for content upload and delivery worldwide. Our service offers unlimited storage, offsite backups on Amazon Glacier, and the option to receive archives via Amazon S3. Our provision of permanent URLs to repository records underscores our commitment to the preservation of information and objects in the long term.

In the remainder of this document, we highlight key services employed to safeguard your content. We encourage you to read carefully and let us know how we can improve our services for the future.

## Persistent URLs

One of the most important aspects of maintaining a reliable repository service is the provision of stable links to your content. Digital Commons provides clean, stable, and easily citable links to your content. These links take the following form: http://your_domain.edu/your_collection/submission number/.

For example, the URL "http://digitalcommons.unl.edu/feralhog/5/" refers to content hosted by the University of Nebraska-Lincoln ("http://digitalcommons.unl.edu") in a collection of papers about feral hogs ("feralhog"), and points to the fifth paper posted to this collection ("5").

We do not change the URLs that point to your content, even if you move a collection within your site's hierarchy. We do, however, provide tools that allow you to manage the content on your webpages or completely remove webpages in the case of mistaken submissions or take-down requests.

## A Robust Infrastructure and Dedicated Staff to Manage It

Bepress hosts 400+ repositories for institutions all over the world. To keep so many sites up and running smoothly we employ well-tested software on a strong system architecture.

**Key points of our system:**

1. All of our production servers are maintained at a high availability colocation facility with multiple backbone connections and backup generators. The facility is secure and requires physical tokens (badges) for access.
2. We maintain failover web, database, and storage servers to continue to serve content in case of failures.
3. Our databases have real-time redundancy that runs continuously, and we take full nightly backups of our entire database. The nightly backups are stored away from our colocation facility in a separate physical location.
4. All of your uploaded files are stored in triplicate in our redundant storage cluster, as well as backed up offsite to a third-party cloud service, Amazon Glacier, that specializes in data archiving and backup.
5. We also send monthly backups of all other data to a third-party archival service, an industry leader in data protection and recovery services. The archival service maintains backup tapes for one year.

**The policies, procedures and staff at bepress are every bit as important as our hardware and software:**

1. We have engineers on call 24/7/365.
2. We automatically monitor all critical system metrics, including server response time for static and dynamic pages and web server load.
3. Our support organization has a detailed plan for supporting subscribers from remote locations in the case of any major disruptions at our home office.

## Privacy and Security

Digital Commons is utilized to publish scholarly, primarily open-access content that is in the public domain and meant for public discovery. Though Digital Commons is not used to store or host sensitive, confidential, personal, or financial information, we do collect some user information to create accounts that allow users to make their scholarly works available.

**Privacy:**

1. The information we collect for a user account is limited to the information readily available on a typical faculty web page or business card. Here are the fields we collect, and only a few are required (those are starred):
   Name*
   Email address*
   Password for the Digital Commons platform*
   University
   Department
   Address
   Phone

2. Digital Commons offers options to notify users and obtain their consent: at the point that they create accounts and at the point that they submit their scholarly content to be published on Digital Commons.
3. We provide granular user account management capability to enforce user security levels within the system.
4. The only method of alteration of any published document in our system is via our administrator interface. All activities and versions are tracked so changes are apparent and recoverable.

**Bepress maintains a best practices approach to security issues. To protect against malicious threats:**

1. We monitor and install necessary patches to software as they are available and firewall all internal resources.
2. We monitor against attacks and suspicious behavior such as automated crawling or other targeted threats to the service and take appropriate action.
3. Our security also includes protection against SQL injection, buffer overflow, XSS, and other attack vectors.
4. Our operations team monitors the CERT advisory and updates the system based on new vulnerabilities.
5. Our backup tapes are encrypted before being sent off-site for archival purposes.
6. We safeguard user information by authenticating user logins through 2048 bit SSL and by integrating with LDAP and CAS subscriber authentication systems when possible.
7. We implement reCAPTCHA on forms to reduce spam.

Of course, none of these services have high value unless your interaction with our software is successful. This insight has given us a strong focus on providing the tools you need most to manage your content and the support required to help you achieve your goals.

**We focus on the technical details so you can focus on your mission—collecting your institution's intellectual assets.**

## bepress Archive

While Digital Commons maintains numerous backups to safeguard your content, we also offer access to an archive of that content with bepress Archive. Clients that maintain an Amazon S3 account can have access to their archive, 24/7. This service also includes automatic checksums and data integrity checks. Clients will have complete control over the contents and administration.

The bepress Archive service offers a solution which is comprehensive, scalable, and immediately available. Once the link to your S3 bucket is established, our archiving service will deposit complete metadata records along with the most recent versions of all posted full-text PDF documents, native files, and supplementary content. The archive's contents are organized in a file/folder hierarchy mirroring the structure of the repository, and are incrementally updated as records and files are added to and revised on the repository. Where applicable, bepress Archive includes content from institutionally-affiliated profiles in SelectedWorks.

## Long-Term Preservation and OAIS Compliance via LOCKSS

In the face of the challenges posed by providing access to digital assets over the long term, the LOCKSS organization (http://www.lockss.org) has developed a very effective tool. The Open Archival Information System-compliant LOCKSS software is designed to harvest and preserve subscription-controlled journal articles, and is well-suited to preserve open access repository content. Bepress has worked with the LOCKSS organization to ensure that Digital Commons repositories are LOCKSS-compliant. Through this compliance, Digital Commons subscribers have joined together to create a Private LOCKSS Network, or PLN, to back up each other's content for long-term preservation.

## Software Testing and Development

The team at bepress uses an iterative methodology for the development of Digital Commons. We have a defined Product Management process which guides feature decisions and delivery. We work in a highly collaborative environment and, as a team, strive to deliver you the highest quality IR on the market. We have implemented rigorous testing procedures:

1. We have multiple test environments to develop and test all features prior to release.
2. We have a dedicated group for quality assurance. They are separate from our development team and they test from the perspective of our users: administrators and site visitors.
3. We test for browser compliance. This includes the most recent versions of: Chrome, Firefox, Safari, and Internet Explorer.
4. We test all Digital Commons features before each software release—not just the new ones. This ensures that new features don't prevent existing features from working properly.
5. After making new features available, our work continues as we track usage and suggestions for improvement to ensure that Digital Commons features remain valuable. Making this effort is an important aspect of our user-focused product development. For example, some of our best features follow a pilot process. We may introduce a feature as a customization for an individual site or sites. After learning from the pilot implementation, we enhance the feature and make it available to all other Digital Commons subscribers as part of our regular releases.

In addition to tracking subscriber requests and improving existing features, we closely follow trends among repository platforms and pay close attention to general web trends. This helps us deliver the contemporary repository services our subscribers expect, in a manner consistent with the other web services they use.

## Format Migration and Emulation

File formats change at a rapid pace. It is likely that we'll be able to open today's Microsoft Word documents with ease next year. It is far less likely that we will be able to open them with ease 20 years from now. This poses an acute threat to institutional repositories, as they are charged with preserving digital assets for long periods of time.

To combat this threat, Digital Commons preserves all objects uploaded to the repository in their original format. This gives us the option to build in capabilities for either emulation (provision of a virtual environment that can display a digital object in its native format), or migration (the creation or transformation from one file format to another). Beyond this, we are committed to making PDFs web-accessible on a permanent basis. If Adobe changes Acrobat to such an extent that older PDFs are no longer readable, we will migrate these objects to ensure they remain readable.

## Scheduled Maintenance

Typical releases include a mix of new features, enhancements to existing features, architectural improvements, and bugfixes.

During a typical maintenance event, we strive to keep impact to a minimum. Maintenance periods are scheduled during periods of low system use, usually late on Friday evenings, Pacific Time (GMT-8). Visitors can view posted submissions and download content. Dynamic pages, including the administrative interface for the service, display a system maintenance message. Our clients receive notice of such events weeks in advance.

## Conclusion

The most important aspect of preservation is an institution's commitment to actively safeguard its content through best practices. Our service has been designed to support this goal by leveraging our experience and technology to provide a stable yet flexible environment for your most important intellectual assets. By working with a large number of diverse institutions we are able to realize economies of scale and pass them along to you through a sustainable pricing structure. For additional information about any of the above, please contact bepress Consulting Services at dc-support@bepress.com or weekdays at (510) 665-1200, option 2, 8:30 a.m.–5:30 p.m. Pacific time.