



29 September 2016

Understanding DOIs – part 1

Jannean Elliott, Manager for Harvesting, Web Services & Data Registration, DOE Office of Scientific and Technical Information (OSTI).



Digital Object Identifiers and DOE's Interagency Data ID Service

Jannean Elliott, Data Services
Department of Energy
Office of Scientific and Technical Information
September 29, 2016
<http://orcid.org/0000-0002-9962-3894>

Digital Object Identifier

- Is one specific type in a class of identifying numbers with certain characteristics. The class name is *Persistent Identifier*.
- Typically, a persistent identifier is not only “persistent” or long-lasting, but “actionable,” meaning a linking action is one of its characteristics.
- The term “persistent identifier” is usually used in the context of the digital world, i.e. the Internet.

- DOI system began in 2000.
- International DOI Foundation is the managing organization.
- DOI name resolution is provided through the Handle System, which was developed by the Corporation for National Research Initiatives (CNRI) and is freely available to anyone.
- A DOI has a distinct format.
- DOI “names” can identify creative works in both electronic and physical forms and at varying levels of detail (granularity).



CrossRef and DataCite are well-known DOI registration agencies.



Allocation agencies obtain DOI prefixes from registration agencies and create/assign DOIs to their own set of clients.

- Each client receives a unique prefix provided by the registration agency with which their allocation agency affiliates.
- Allocation agency ensures each DOI gets a unique suffix.

Then there are DOI
formats with
“intelligence”

10.18942/178867

or

<http://dx.doi.org/10.18942/178867>

- 10.1038/ncomms11735: Infix is name of journal, *Nature Communications*
- 10.3334/ORNLDAAC/849: Infix is name of repository, ORNL Distributed Active Archive Center for Biogeochemical Dynamics
- 10.14291/tccon.ggg2014.garmisch01R0/1149299: Infix is the project, the year, the collection site, the “site within the site,” and the revision number. Total Column Carbon Observing Network (TCCON) Site: Garmisch

Who Cares?

Klein, M., Van de Sompel, H., Sanderson, R., Shankar, H., Balakireva, L., Zhou K., and Tobin, R. (2014) Scholarly Context **Not** Found: One in Five Articles Suffers from Reference Rot. PLoS ONE, 9(12): e115253. doi: [10.1371/journal.pone.0115253](https://doi.org/10.1371/journal.pone.0115253)



Herbert Van de Sompel, an information scientist at the Los Alamos National

Laboratory Research Library in New Mexico, quantified the alarming extent of this 'link rot' and 'content drift' (together, 'reference rot') in a paper published last December. With a group of researchers under in the Hiberlink project, he analysed more than 1 million 'web-at-large' links (defined as those beginning with 'http://' that point to sites other than research articles) in some 3.5 million articles published between 1997 and 2012. The team found that in articles from 2012 ... 22% of hyperlinks in papers from Elsevier journals were rotten (the proportion rises in older articles), and overall some 75% of links were not cached on any Internet archiving site within two weeks of the article's publication date..... Some four-fifths of 2012 papers that contained a link suffered from reference rot, the team found — that is, at least one reference to web-at-large content was either dead or not archived. Van de Sompel terms the situation **“rather dramatic”**.

NATURE | TOOLBOX Sharing The trouble with reference rot.
Jeffrey M. Perkel 04 May 2015 Vol 521, issue 7550



Why Did DOE/OSTI Choose Digital Object Identifiers (DOIs)?

Pre-existing global recognition and acceptance in the world of scholarly publications

Standardization through the backing of the International DOI Foundation (IDF) and the National Information Standards Organization (NISO)

OSTI's prior experience with CrossRef

... and then along came DataCite



- A global non-profit comprised of local institutions focused on improving the scholarly infrastructure around data
- Officially launched December 2009
- Executive Board, small full-time staff, members, Managing Agent TIB.
- Members may act as Allocating Agents and provide DOIs to clients.
- Maintains registry of data DOIs and the re3data.org registry of research data repositories.

Three U.S. Members:
(until FY16. Now growing)

California Digital Library
Purdue University

DOE/OSTI

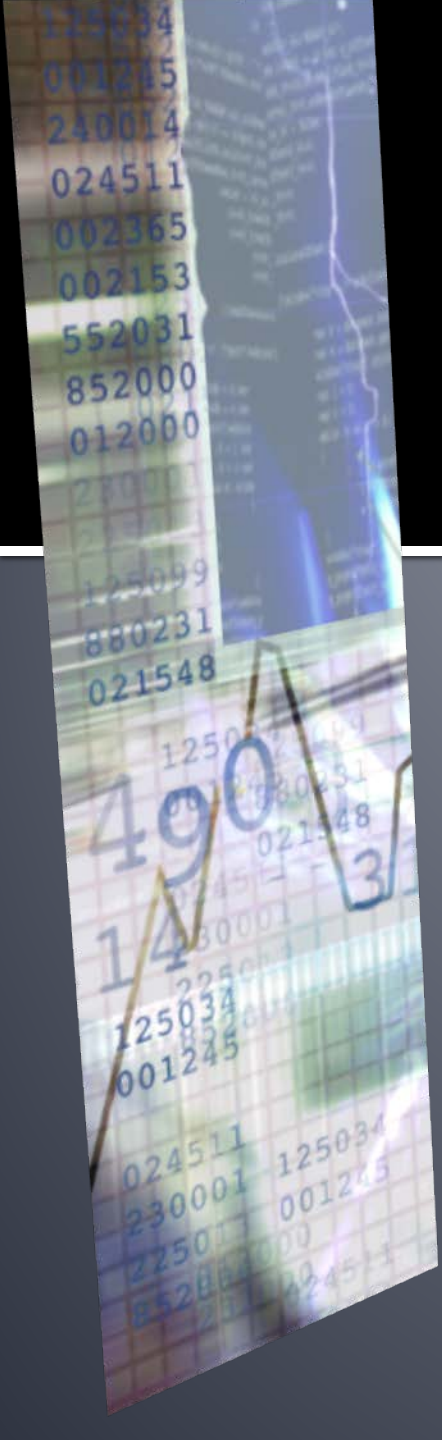
- OSTI serves DOE with the Data ID Service, free of charge.
- Serves other U.S. Federal Agencies on a cost-recovery basis.
- Full membership approved December 2010.
- First DOI assigned and minted August 2011.

DataCite Members: 2010 - 15 members in 10 countries; 2014 - 22 members in 19 countries; now growing rapidly

- 📍 National Research Council of Thailand
- 📍 Hungarian Academy of Sciences
- 📍 South African Environmental Observers
- 📍 European Organisation for Nuclear Research
- 📍 University of Tartu
- 📍 Japan Link Center - JaLC

- 📍 Australian National Data Service
- 📍 Canada Institute for Scientific and Technical Information
- 📍 Technical Information Center of Denmark
- 📍 Institute for Scientific and Technical Information
- 📍 German National Library of Science and Technology
- 📍 German National Library of Medicine
- 📍 GESIS - Leibniz Institute for the Social Sciences
- 📍 CRUI - Conference of Italian Universities
- 📍 TU Delft Library
- 📍 The Swedish National Data Service
- 📍 ETH Zurich
- 📍 The British Library
- 📍 German National Library of Economics
- 📍 California Digital Library
- 📍 Office of Scientific and Technical Information
- 📍 Purdue University Library





DOE's *Interagency Data ID (IAD) Service* is the “umbrella” name for the dataset submission process, the DOI assignment process, and the DOI registration process for those datasets.

Supporting data citation to facilitate:

Long-term discovery of research objects

Access through stable linkage and identification

Better attribution for researchers and data producers

Better budget and impact tracking for sponsors

Easier reuse and verification of data

New data services of the future

The IAD Process (in general)

- Agency determines that important datasets exist which need to be announced for visibility and assigned DOIs.
- Agency POC contacts IAD; conference calls begin, and Interagency Agreement paperwork is completed.
- Metadata is submitted through offline formatting and uploaded via IAD web page or via an automated 241.6 web service/API.
- IAD obtains unique DataCite prefix for new data client and establishes accounts/credentials.
- Testing, review, and feedback cycles are completed.
- Data client and IAD POC agree all is ready to move to production.
- OSTI provides year-round support, help with updates, etc.

What *really* makes a DOI special ?

Answer (IMHO):

The implied commitment it represents

- Questions on granularity of DOI assignment are carefully discussed.
- Product/resource types, metadata, business rules
- Landing pages

Submitter Responsibilities for Dataset Registration

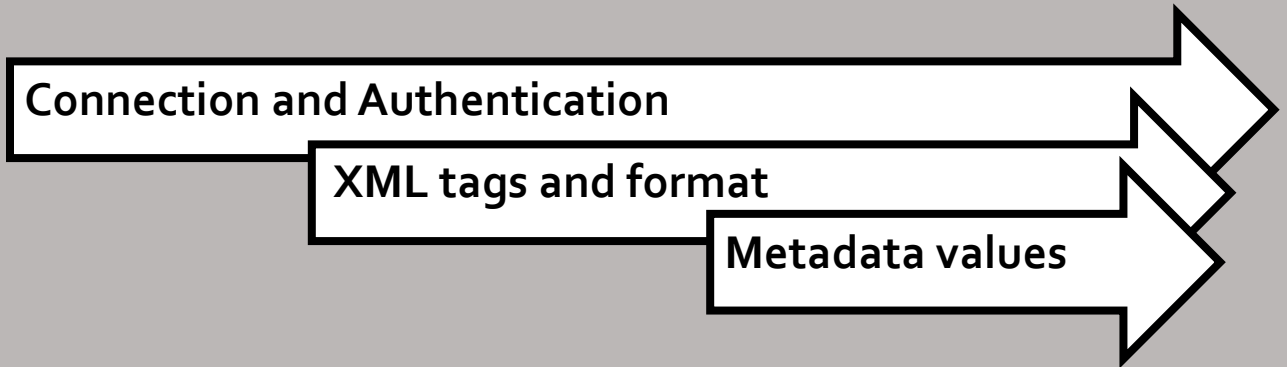
- Provide, at minimum, the required metadata to enable basic, bibliographic citation
- Have authority (though not necessarily ownership) to make public the assets for which they wish to assign the DOIs.
- Guarantee the persistence of registered data
 - By ensuring that data will be stored and managed such that access and usability are provided indefinitely.
 - By maintaining and updating as necessary all URLs associated with the DOI
- Establish and maintain appropriate landing pages for registered DOIs

Subject expertise and knowledge of how the intended audience will expect to look for and use the data is the key to successfully defining content and dataset boundaries.

TEST Phase Begins

Development

- ▶ During TEST phase, client is receiving TEST DOIs with DataCite's TEST DOI prefix. TEST metadata goes to DataCite's MDS but not into the Registry. TEST DOIs will not resolve.



A cycle of review and feedback

**Then the landing
page questions
really begin to
come!**



The Landing Page

The landing page, which must always be publicly accessible, provides context for the dataset. It may contain one or more of the following:


- Information on how a full, formal citation of the data should appear
- Access information, including a notice of temporary restriction, if a dataset is not yet ready for public access
- Links to software or toolkits needed to open, download, or analyze the data
- A listing of the individual data files that may be included in the registered dataset (with links to them)
- Update and version information
- Contact information.



The Landing Page

Geography

[Enlarge Map](#)



Keyword

Federal Agency
Department of Energy (1)

Data Originator
NGEE Arctic (1)

Scientist
Alistair Rogers (arogers@bnl.gov) (1)

< 1 > displaying 1 to 1 of 1

Results: 10 Sort By: Relevance

Plant Physiology Data, Barrow, Alaska, 2013

NGEE Public Dataset

Data Source: NGEE Arctic

Begin Date: 07/09/2013 **End Date:** 08/11/2013

Federal Agency: Department of Energy

Survey measurements of photosynthetic rate and stomatal conductance together with carbon dioxide concentration, temperature, PAR, and relative humidity for 8 species on the BEO. <http://dx.doi.org/10.5> [more](#)

Get Data

View Metadata

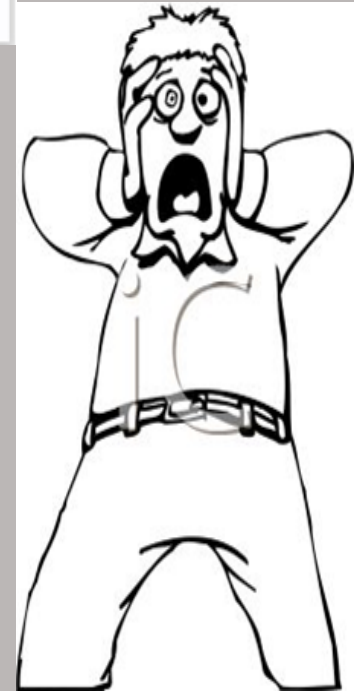
< 1 > displaying 1 to 1 of 1

“Landing” in a searchable data catalog
Next Generation Ecosystem Experiments – Arctic



< Useful when data must be retired

**DataCite
Does
Not
Delete !**



Creating an “Intelligent” DOI

Production

Production

Components:

DataCite
Schema

XML Transfer
Streams

DataCite

Metadata Store

Production

Phases:

Translation and
transfer

Validation

Minting

Availability

- Allows a data client to provide a character string in the submitted record to be added to the DOI that the IAD Service assigns.
- XML tag name is <doi_infix> ; the value put in this tag by the client can be letters, numbers, or a combination. The infix value might be a project name, a category type, a geographic region, or anything that will add “intelligence,” i.e. specific meaning for a specific set of users.
- IAD takes the infix value provided and inserts it between the unique prefix for the client and the unique OSTI ID that forms the prefix.

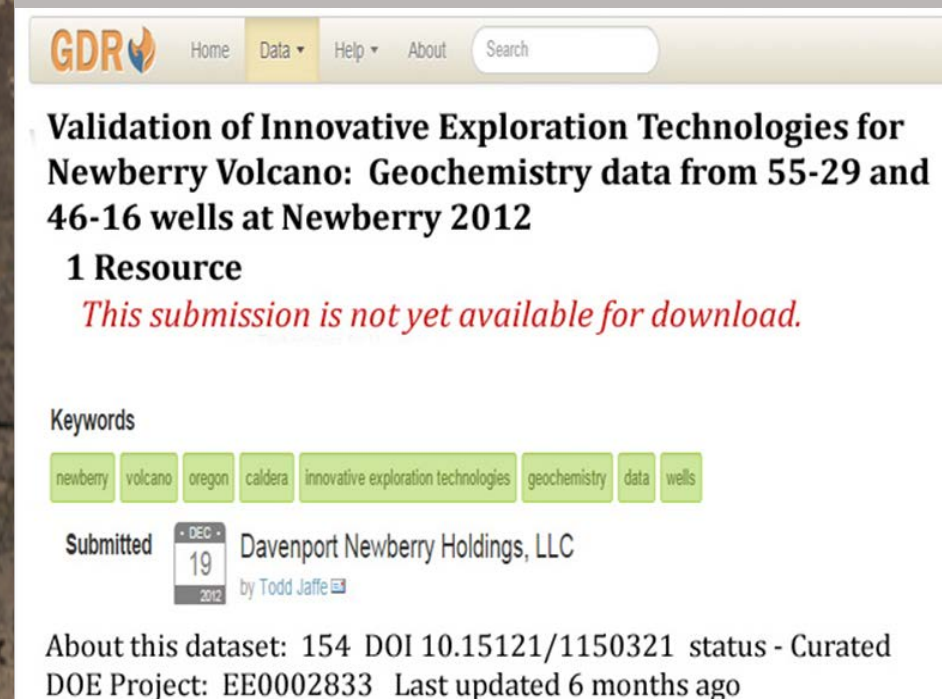
DOI prefix/DOI infix/DOI suffix

prefix from DataCite/infix from submitter/suffix from OSTI ID

10.19597/myprojectname/1105143

Reserving a DOI

This GDR dataset was a straightforward, normal submission because the landing page, with good information, was publicly available at the time of submission. The DOI did not have to be reserved.



The screenshot shows the GDR (Geological Data Repository) landing page for a dataset. The page title is "Validation of Innovative Exploration Technologies for Newberry Volcano: Geochemistry data from 55-29 and 46-16 wells at Newberry 2012". It indicates "1 Resource" and states "This submission is not yet available for download." The page includes a navigation bar with "Home", "Data", "Help", and "About" links, and a search box. Below the title, there are keywords: "newberry", "volcano", "oregon", "caldera", "innovative exploration technologies", "geochemistry", "data", and "wells". The submission information shows it was submitted on December 19, 2012, by Todd Jaffe, on behalf of Davenport Newberry Holdings, LLC. At the bottom, it provides the DOI "10.15121/1150321" and notes the dataset is curated by the DOE Project "EE0002833", last updated 6 months ago.

GDR Home Data Help About Search

Validation of Innovative Exploration Technologies for Newberry Volcano: Geochemistry data from 55-29 and 46-16 wells at Newberry 2012

1 Resource

This submission is not yet available for download.

Keywords

newberry volcano oregon caldera innovative exploration technologies geochemistry data wells

Submitted **DEC 19 2012** Davenport Newberry Holdings, LLC by Todd Jaffe

About this dataset: 154 DOI 10.15121/1150321 status - Curated DOE Project: EE0002833 Last updated 6 months ago



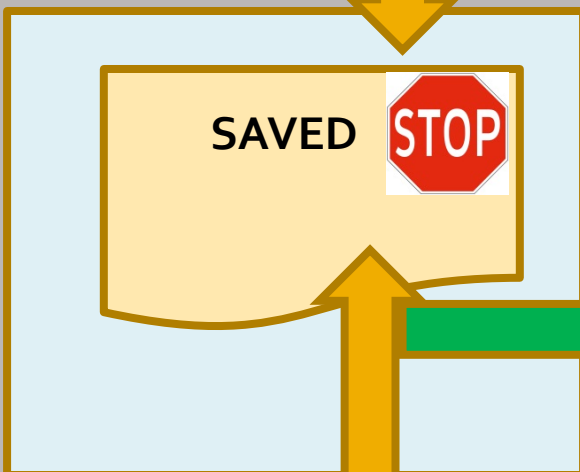
Core scan2 Kimama Well

Sometimes a researcher needs to know what the DOI for her dataset will be even before she knows where its landing page will be posted. Reserving a DOI in IAD supports this need. A DOI for a particular dataset is assigned, but IAD ensures that the DOI does not “travel” to DataCite before the researcher is ready for it to do so. If the DOI “travels” before there is a valid URL, DataCite will not mint. The DOI would not be registered and cannot function as a live link.

First submission

<set_reserved/>

January



Released Record

June

<osti_id>
<site_url>

Update submission

DOI Stream

Metadata Stream



When OSTI and the client agree they are mutually ready, OSTI sets up the web service account in production IAD.

- ▶ OSTI’s server provides immediate response to client after each submission, including email to client-specified distribution.

DOI	Is Active	Is Ref Quality	Updated	Minted	Latest Metadata Version	
10.5439/1171947	true	false	2015-03-21 02:16 UTC	2015-03-21 02:16 UTC	0 (2015-03-21 02:16:22.0)	
10.5439/1171946	true	false	2015-03-21 02:16 UTC	2015-03-21 02:16 UTC	0 (2015-03-21 02:16:21.0)	
10.5439/1171945	true	false	2015-03-21 02:16 UTC	2015-03-21 02:16 UTC	0 (2015-03-21 02:16:20.0)	
10.5439/1171944	true	false	2015-03-21 02:16 UTC	2015-03-21 02:16 UTC	0 (2015-03-21 02:16:18.0)	
10.5439/1171943	true	false	2015-03-21 02:16 UTC	2015-03-21 02:16 UTC	0 (2015-03-21 02:16:17.0)	
10.5439/1171942	true	false	2015-03-21 02:16 UTC	2015-03-21 02:16 UTC	0 (2015-03-21 02:16:16.0)	
10.5439/1164928	true	false	2015-03-21 02:15 UTC	2015-03-21 02:15 UTC	0 (2015-03-21 02:15:51.0)	
10.5439/1171962	true	false	2015-03-07 03:30 UTC	2015-03-07 03:30 UTC	0 (2015-03-07 03:30:00.0)	
10.5439/1169528	true	false	2015-02-20 03:19 UTC	2015-02-10 03:16 UTC	1 (2015-02-20 03:19:07.0)	
10.5439/1169527	true	false	2015-02-10 03:16 UTC	2015-02-10 03:16 UTC	0 (2015-02-10 03:16:48.0)	
10.5439/1169526	true	false	2015-02-10 03:16 UTC	2015-02-10 03:16 UTC	0 (2015-02-10 03:16:46.0)	
10.5439/1169525	true	false	2015-02-10 03:16 UTC	2015-02-10 03:16 UTC	0 (2015-02-10 03:16:44.0)	
10.5439/1169523	true	false	2015-02-10 03:16 UTC	2015-02-10 03:16 UTC	0 (2015-02-10 03:16:43.0)	
10.5439/1169522	true	false	2015-02-10 03:16 UTC	2015-02-10 03:16 UTC	0 (2015-02-10 03:16:41.0)	

▼ Show Dataset

DOI: [10.11577/1172615](#)

DOI latency: Be aware that it can take up to 24 hours until a DOI update is globally known. New DOIs should be resolvable after about 5 minutes.

Minted: 2015-03-17 02:25 UTC

Updated: 2015-03-17 02:25 UTC

Is Active: true

- ▶ The metadata record is released and travels to “DataCite.” New DOIs “mint” immediately. Updates may not go “live” for several hours.

Interagency Data (IAD) Web Service



Office of Scientific and Technical Information

Home

Log In

The Office of Scientific and Technical Information (OSTI) became a member and an allocation agency for DataCite in 2011. OSTI assigns Digital Object Identifiers (DOIs) to Department of Energy datasets and registers those DOIs with DataCite. The Interagency Data ID Service (IAD), developed and operated by OSTI, provides this same service to other U.S. Federal Agencies desiring to make their data available for citation and discovery long into the future.

[Access user instructions and documentation for IAD.](#) Instructions for formatting an XML record or file to upload via this web page and instructions for automated submission to IAD's web service/API are included.

Select an XML file to upload to the Interagency Data ID Service.

Choose File No file chosen

Upload XML

Contact

To express interest in using the IAD, or for credentials or technical questions:

IAD Services
IAD@osti.gov



NIMH Data Archive

National Database for Autism Research

*The Networking and Information Technology
Research and Development (NITRD) Program*



Topics



Agricultural Products



Agroecosystems & Environment



Animals & Livestock



Bioenergy



Food & Nutrition



Genomics & Genetics



Maps & Multimedia



Plants & Crops

Highlighted Datasets



Nutrient and herbicide concentrations, loads, and daily discharge data for caves in the Goodwater Creek Experimental Watershed, a Long-Term...



USDA-ARS carries out water productivity field trials for maize in Colorado, analyzing irrigation, precipitation, soil water storage....



The Baylor College of Medicine recently sequenced and annotated the *Oncopeltus fasciatus* genome as part of the 15k pilot project. The...



United States Department of Transportation

OFFICE OF THE ASSISTANT SECRETARY FOR RESEARCH AND TECHNOLOGY
National Transportation Library



Interested in becoming an IAD client?

Jannean Elliott

elliottj@osti.gov

865-576-6784

Crystal Sherline

sherlinec@osti.gov

865-576-6597

<https://www.osti.gov/iad/>

<http://www.osti.gov/>

Interagency Data (IAD) Web Service

U.S. DEPARTMENT OF **ENERGY**
Office of Scientific and Technical Information

Home Log In

The Office of Scientific and Technical Information (OSTI) became a member and an allocation agency for DataCite in 2011. OSTI assigns Digital Object Identifiers (DOIs) to Department of Energy datasets and registers those DOIs with DataCite. The Interagency Data ID Service (IAD), developed and operated by OSTI, provides this same service to other U.S. Federal Agencies desiring to make their data available for citation and discovery long into the future.

Access user instructions and documentation for IAD. Instructions for formatting an XML record or file to upload via this web page and instructions for automated submission to IAD's web service/API are included.

Contact
To express interest in using the IAD, or for credentials or technical questions:
IAD Services
IAD@osti.gov

Select an XML file to upload to the Interagency Data ID Service.
Choose File No file chosen

U.S. DEPARTMENT OF **ENERGY** | Office of Science | Office of Scientific and Technical Information

OSTI.gov
Speeding access to science information from DOE and beyond

HOME ABOUT OSTI SCIENCE SEARCH TOOLS DOE PAGES^{Beta} / PUBLIC ACCESS COMMUNICATIONS DOE STI PROGRAM

Find DOE R&D Results **GO**

SciTech Connect

DOE Scientific and Technical Information...and more

SciTech Connect
DOE science, technology, and engineering research information.

DOE PAGES^{Beta}
Public Access Gateway for Energy & Science
Public access to DOE-funded peer-reviewed journal articles and accepted manuscripts.

For additional information, see the OSTI Catalogue of Collections.

The Grand Compromise of U.S. Public Access Programs: Going Green