U.S. DEPARTMENT OF ENERGY | Office of Science

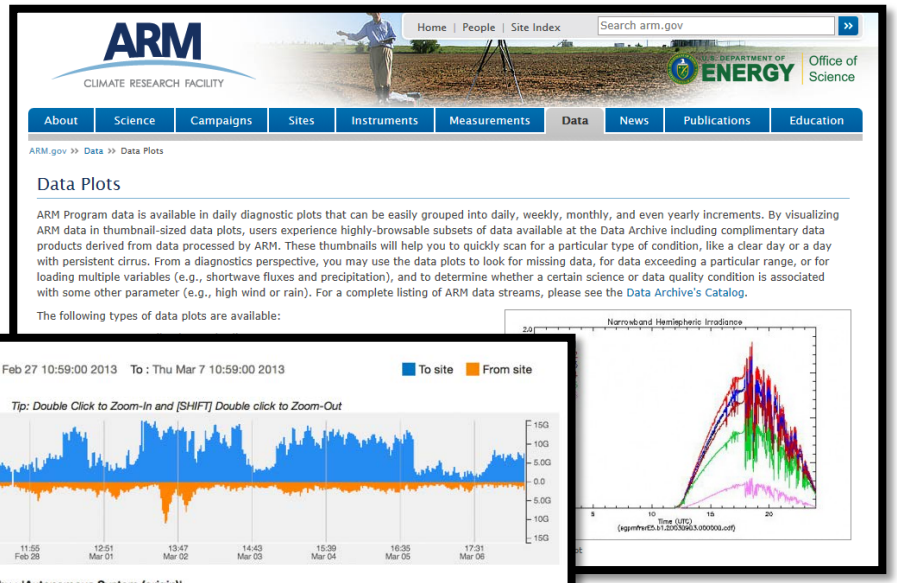
Perspectives from DOE

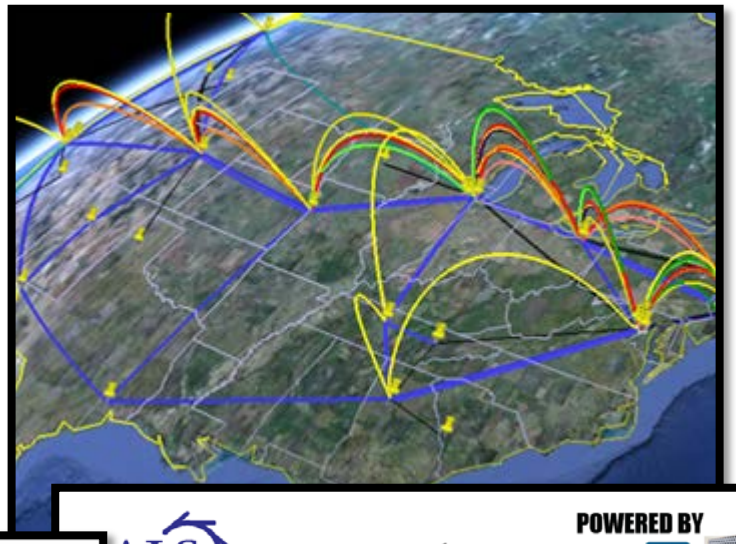
Joint Facilities User Forum on Data-Intensive Science

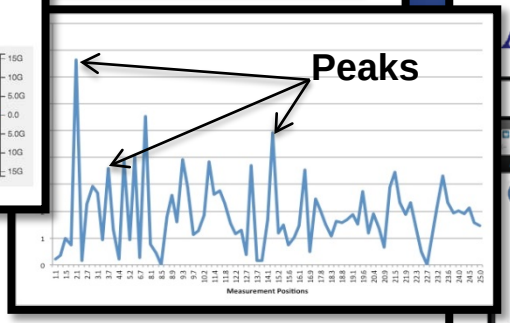
16 June, 2014

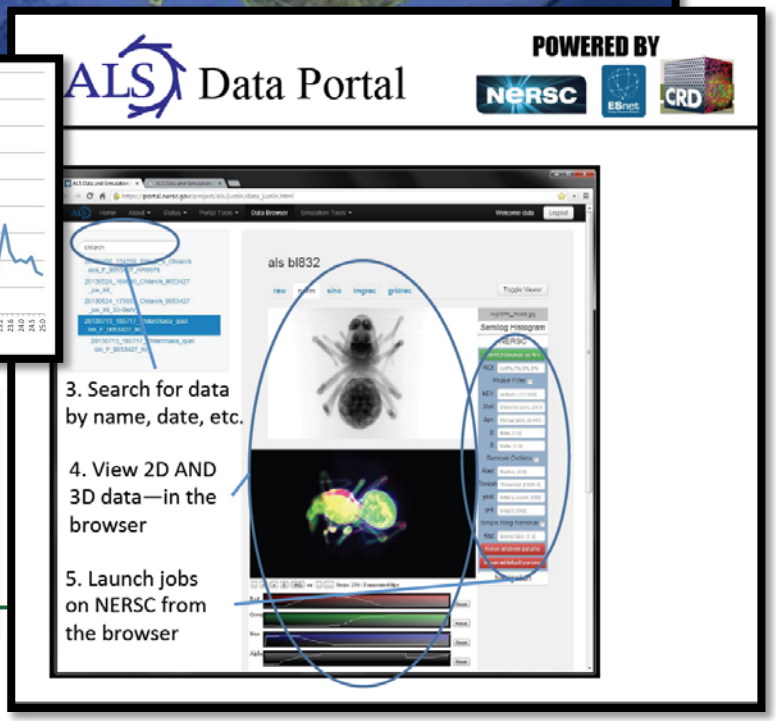
Laura Biven, PhD
Senior Science and Technology Advisor
Office of the Deputy Director for Science Programs (SC-2)
Laura.Biven@science.doe.gov

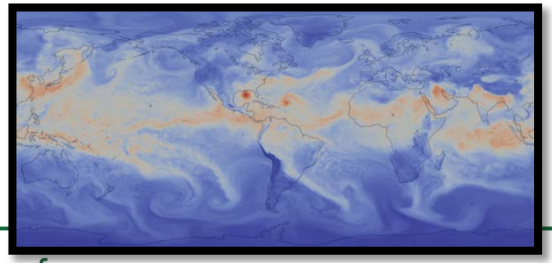
Perspectives on Data











U.S. DEPARTMENT OF ENERGY | Office of Science

Perspectives on Data (from Washington)

OSTP Memo

EXECUTIVE OFFICE OF THE PRESIDENT
OFFICE OF SCIENCE AND TECHNOLOGY POLICY
WASHINGTON, D.C. 20502

February 22, 2013

MEMORANDUM FOR THE HEADS OF EXECUTIVE DEPARTMENTS AND AGENCIES

FROM: John P. Holdren
Director

SUBJECT: Increasing Access to the Results of Federally Funded Scientific Research

1. Policy Principles

The Administration is committed to ensuring that, to the greatest extent and with the fewest constraints possible and consistent with law and the objectives set out below, the direct results of federally funded scientific research are made available to and useful for the public, industry, and the scientific community. Such results include peer-reviewed publications and digital data.

Scientific research supported by the Federal Government catalyzes innovative breakthroughs that drive our economy. The results of that research become the grist for new insights and are assets for progress in areas such as health, energy, the environment, agriculture, and national security.

Access to digital data sets resulting from federally funded research allows companies to focus

DOE Proposed Response

Data

- **SC Statement on Digital Data Management**

Publication

- **Public Access Gateway for Energy Sciences (PAGES)**

DOE Draft "Public Access Plan"
has been submitted to OSTP

Summary: Office of Science Statement on Digital Data Management

Principles

- Effective data management has the potential to increase the pace of scientific discovery and promote more efficient and effective use of government funding and resources. Data management planning should be an integral part of research planning.

- Sharing and preserving data are central to protecting the integrity of science by facilitating validation of results and to advancing science by broadening the value of research data to disciplines other than the originating one and to society at large.

- Not all data need to be shared or preserved. The costs and benefits of doing so should be considered in data management planning.

Summary: Office of Science Statement on Digital Data Management

Requirements (1 of 3)

- To integrate data management planning into the overall research plan, **all proposals submitted to the Office of Science for research funding are required to include a Data Management Plan (DMP)** that describes how data generated through the course of the proposed research will be shared and preserved or explains why data sharing and/or preservation are not possible or scientifically appropriate. At a minimum, DMPs must describe how data sharing and preservation will enable validation of results, or how results could be validated if data are not shared or preserved.

DMPs will be reviewed as part of the overall Office of Science research proposal merit review process.

Summary: Office of Science Statement on Digital Data Management

Requirements (2 of 3)

- **DMPs must provide a plan for making all research data displayed in publications resulting from the proposed research digitally accessible at the time of publication**. This includes data that are displayed in charts, figures, images, etc. This requirement could be met by including the data as supplementary information to the published article, or through other means. The published article should indicate how these data can be accessed.

Summary: Office of Science Statement on Digital Data Management

Requirements (3 of 3)

- In determining the resources needed for data management, **researchers that plan to work at an Office of Science User Facility as part of the proposed research should consult the published data policy of that facility** and reference it in the DMP. DMPs that explicitly or implicitly commit data management resources at a facility beyond what is conventionally made available to approved users should be accompanied by written approval from that facility.

Definitions

Digital Research Data:

The term *digital data* encompasses a wide variety of information stored in digital form including: experimental, observational, and simulation data; codes, software and algorithms; text; numeric information; images; video; audio; and associated metadata. It also encompasses information in a variety of different forms including raw, processed, and analyzed data, published and archived data.

This statement focuses on *digital research data,* which are *research data* that can be stored digitally and accessed electronically. OMB Circular A110 defines *research data* as follows:

> "Research data is defined as the recorded factual material commonly accepted in the scientific community as necessary to validate research findings, but not any of the following: preliminary analyses, drafts of scientific papers, plans for future research, peer reviews, or communications with colleagues. This 'recorded' material excludes physical objects (e.g., laboratory samples). Research data also do not include:
>
> (A) Trade secrets, commercial information, materials necessary to be held confidential by a researcher until they are published, or similar information which is protected under law; and
>
> (B) Personnel and medical information and similar information the disclosure of which would constitute a clearly unwarranted invasion of personal privacy, such as information that could be used to identify a particular person in a research study."

Office of Science Statement on Digital Data Management

- **Requirements will apply to all proposals for research funding regardless of institution but NOT to applications for time on user facilities**

- **Requirements will apply to proposals submitted in response to all Office of Science research solicitations and invitations for new, renewal, and some supplemental funding issued on or after DATE TBD**

"Objectives for Public Access to Scientific Publications

"To the extent feasible and consistent with law; agency mission; resource constraints; U.S. national, homeland, and economic security… the results of unclassified research that are published in peer-reviewed publications directly arising from Federal funding should be stored for long-term preservation and publicly accessible to search, retrieve, and analyze in ways that maximize the impact and accountability of the Federal research investment."

DOE's Draft Public Access Plan

The Department proposes a model for ensuring public access to scholarly publications resulting from DOE funding that provides the public with access to the **best available version** of the article **within 12 months from publication**.

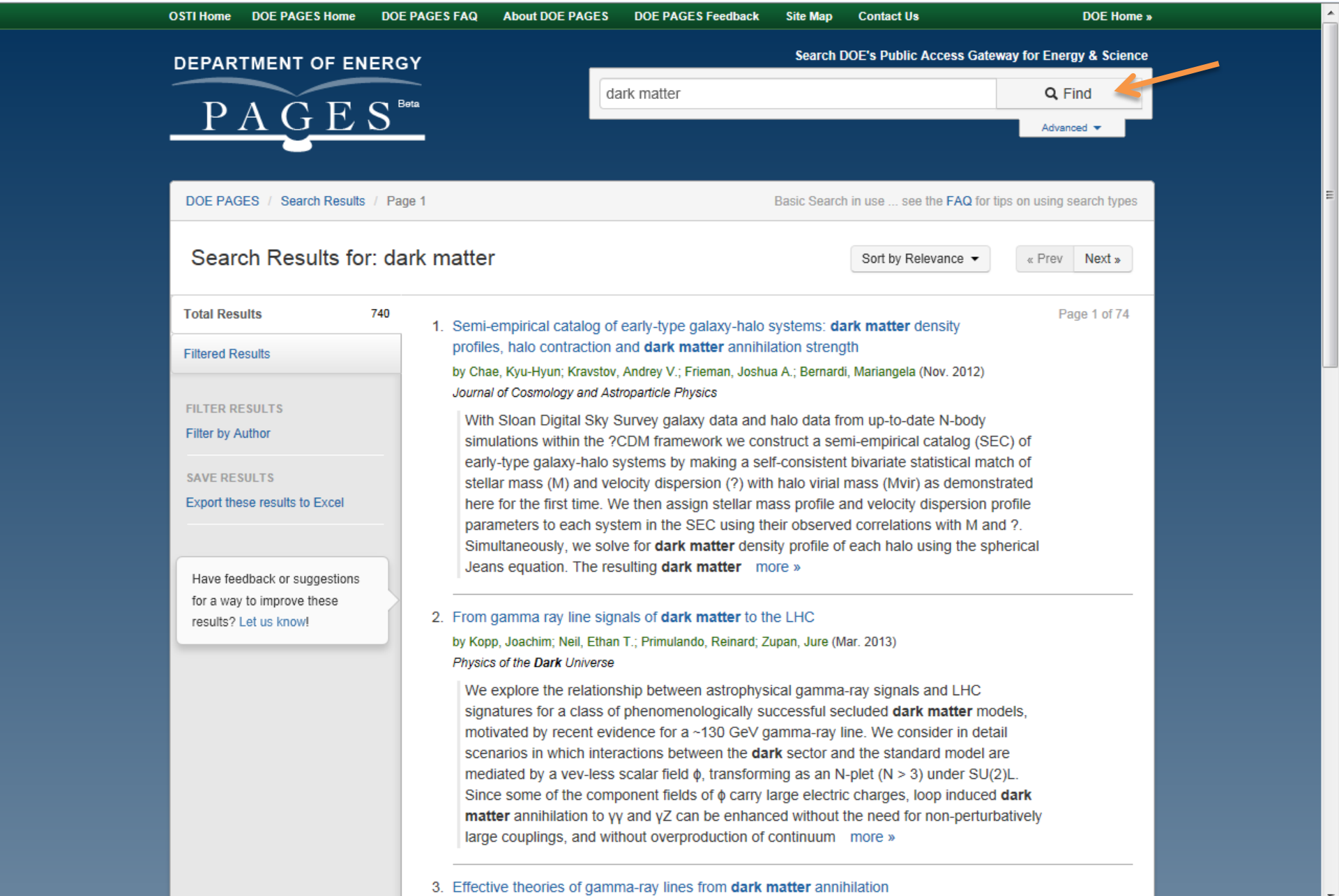
Best available version:

1. Version of Record (VoR) published and hosted by the publisher
2. Accepted manuscript hosted by the publisher, a third party repository (e.g. Lab, arXiv, institutional repository), **or by DOE Office of Scientific and Technical Information (OSTI)**.

Public Access Gateway for Energy and Science (PAGES)

- The Department proposes to host a public-facing portal and search interface tool to enhance discoverability, **PAGES**.
  - For DOE funded publications, PAGES will provide metadata, abstracts, and links to best available version.
  - OSTI will maintain a dark archive of all accepted manuscripts, providing access only to those that are not publically available elsewhere 12 months after publication.
- **There will be new requirements for researchers to submit metadata and accepted manuscripts to OSTI.**
- PAGES is developed and maintained by OSTI for DOE.
- PAGES exists but is not yet publically accessible.

PAGES Search Results



PAGES Item

DEPARTMENT OF ENERGY

PAGES Beta

Search DOE's Public Access Gateway for Energy & Science

dark matter | Find | Advanced

Semi-empirical catalog of early-type galaxy-halo systems: dark matter density profiles, halo contraction and dark matter annihilation strength

Citation Details

With Sloan Digital Sky Survey galaxy data and halo data from up-to-date N-body simulations within the ?CDM framework we construct a semi-empirical catalog (SEC) of early-type galaxy-halo systems by making a self-consistent bivariate statistical match of stellar mass (M) and velocity dispersion (?) with halo virial mass (Mvir) as demonstrated here for the first time. We then assign stellar mass profile and velocity dispersion profile parameters to each system in the SEC using their observed correlations with M and ?. Simultaneously, we solve for **dark matter** density profile of each halo using the spherical Jeans equation. The resulting **dark matter** density profiles deviate in general from the dissipationless profile of Navarro-Frenk-White or Einasto and their mean inner density slope and concentration vary systematically with Mvir. Statistical tests of the distribution of profiles at fixed Mvir rule out the null hypothesis that it follows the distribution predicted by dissipationless N-body simulations for Mvir1013.5 ? 14.5 M?. These **dark matter** profiles imply that **dark matter** density is, on average, enhanced significantly in the inner region of halos with Mvir1013.5 ? 14.5 M? supporting halo contraction. The main characteristics of halo contraction are: (1) the mean **dark matter** density within the effective radius more »

Authors:	Chae, Kyu-Hyun; Kravstov, Andrey V.; Frieman, Joshua A.; Bernardi, Mariangela
Publication Date:	2012-11-05
OSTI Identifier:	OSTI ID: 1076131
Type:	Published Article
Journal Name:	Journal of Cosmology and Astroparticle Physics

Free Publicly Accessible Full Text

Publisher's Version of Record at http://dx.doi.org/10.1088/1475-7516/2012/11/004

Have feedback or suggestions for a way to improve these results? Let us know!

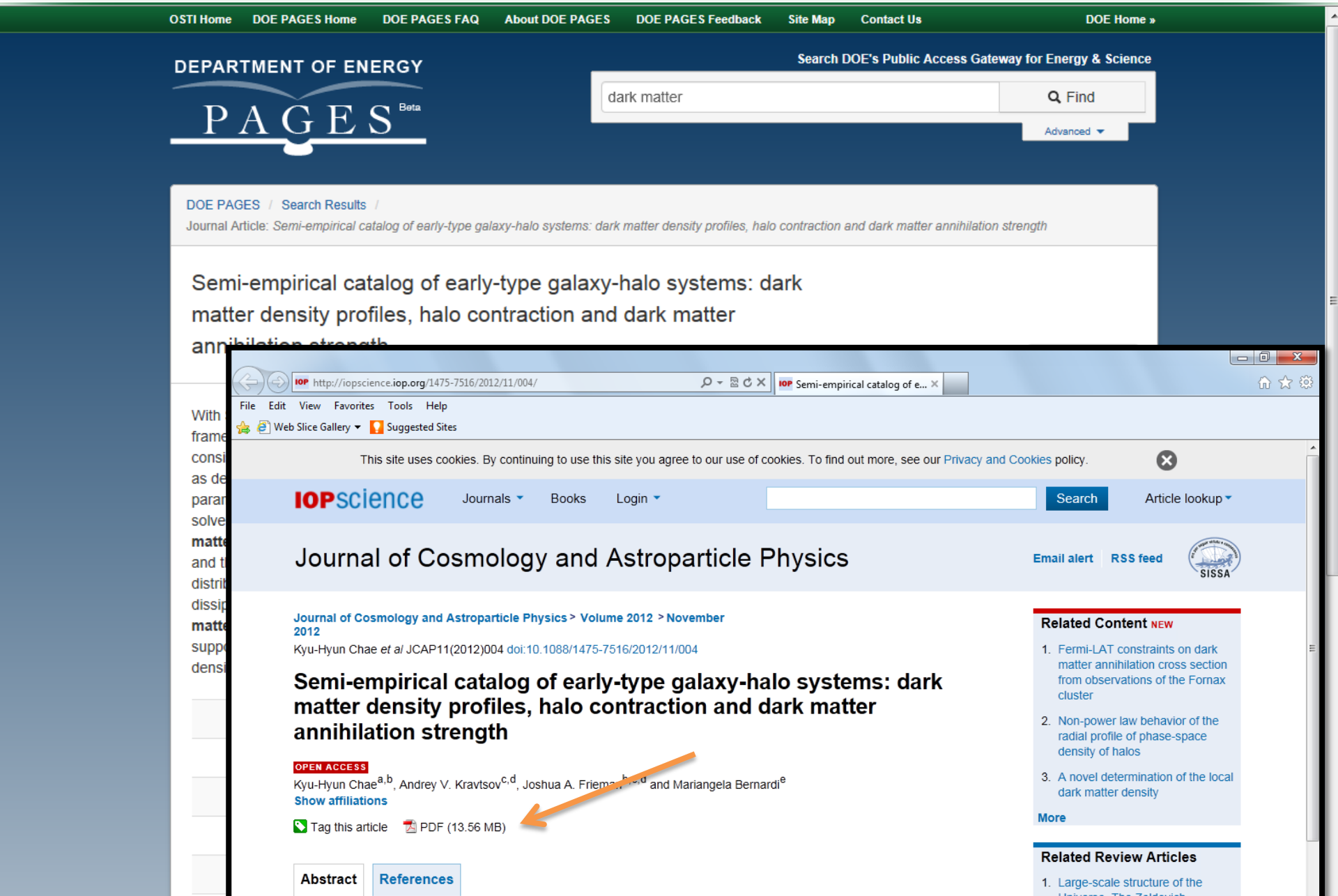
CITATION FORMATS

MLA
APA
Chicago
Bibtex

EXPORT METADATA

EndNote
Excel

PAGES Link to Publisher VoR



PAGES Item within 12 Month Embargo

DEPARTMENT OF ENERGY

PAGES Beta

Search DOE's Public Access Gateway for Energy & Science

1103297 Find

Advanced

This content will become publicly available on Nov. 1, 2014

Comparative study of nuclear effects in polarized electron scattering from 3He

Citation Details

We present a detailed analysis of nuclear effects in inclusive electron scattering from polarized {sup 3}He nuclei for polarization asymmetries, structure functions and their moments, both in the nucleon resonance and deep-inelastic regions. We compare the results of calculations within the weak binding approximation at finite Q^2 with the effective polarization ansatz often used in experimental data analyses, and explore the impact of \Delta components in the nuclear wave function and nucleon off-shell corrections on extractions of the free neutron structure. Using the same framework we also make predictions for the Q^2 dependence of quasielastic scattering from polarized {sup 3}He, data on which can be used to constrain the spin-dependent nuclear smearing functions in {sup 3}He.

Authors:	Ethier, Jacob James; Melnitchouk, Wally [JLAB]

Publication Date:	2013-11-01
OSTI Identifier:	OSTI ID: **1103297**
Report Number(s):	JLAB-THY-13-1770; DOE/OR/23177-2696
DOE Contract Number:	AC05-06OR23177
Type:	Accepted Manuscript
Journal Name:	Phys. Rev. C
Volume:	88

Free Publicly Accessible Full Text

This content will become publicly available on Nov. 01, 2014

Publisher's Version of Record

http://dx.doi.org/10.1103/PhysRevC.88.054001

Have feedback or suggestions for a way to improve these results? Let us know!

CITATION FORMATS

MLA
APA
Chicago
Bibtex

PAGES Item hosted by OSTI

DEPARTMENT OF ENERGY

PAGES Beta

Search DOE's Public Access Gateway for Energy & Science

1095010 Find

Advanced

Comparison of high-pressure CO{sub 2} sorption isotherms on Eastern and Western US coals

Citation Details

Accurate estimation of carbon dioxide (CO{sub 2}) sorption capacity of coal is important for planning the CO {sub 2} sequestration efforts. In this work, we investigated sorption and swelling behavior of several Eastern and Western US coal samples from the Central Appalachian Basin and from San Juan Basin. The CO{sub 2} sorption isotherms have been completed at 55{degrees}C for as received and dried samples. The role of mineral components in coal, the coal swelling, the effects of temperature and moisture, and the error propagation have been analyzed. Changes in void volume due to dewatering and other factors such as temporary caging of carbon dioxide molecules in coal matrix were identified among the main factors affecting accuracy of the carbon dioxide sorption isotherms. The (helium) void volume in the sample cells was measured before and after the sorption isotherm experiments and was used to build the volume-corrected data plots.

Authors:	Romanov, V.; Hur, T.-B.; Fazio, J.; Howard, B
Publication Date:	2012-10-01
OSTI Identifier:	OSTI ID: **1095010**
Report Number(s):	NETL-PUB-348
DOE Contract Number:	FE0004000
Type:	Accepted Manuscript
Journal Name:	29th Int. Pittsburgh Coal Conf., Pittsburgh
Research Org:	National Energy Technology Laboratory - In-house Research

Free Publicly Accessible Full Text

Accepted Manuscript

0.22 MB

Publisher's Version of Record

http://dx.doi.org/10.978.189097 7/290

Have feedback or suggestions for a way to improve these results? Let us know!

CITATION FORMATS

MLA

APA

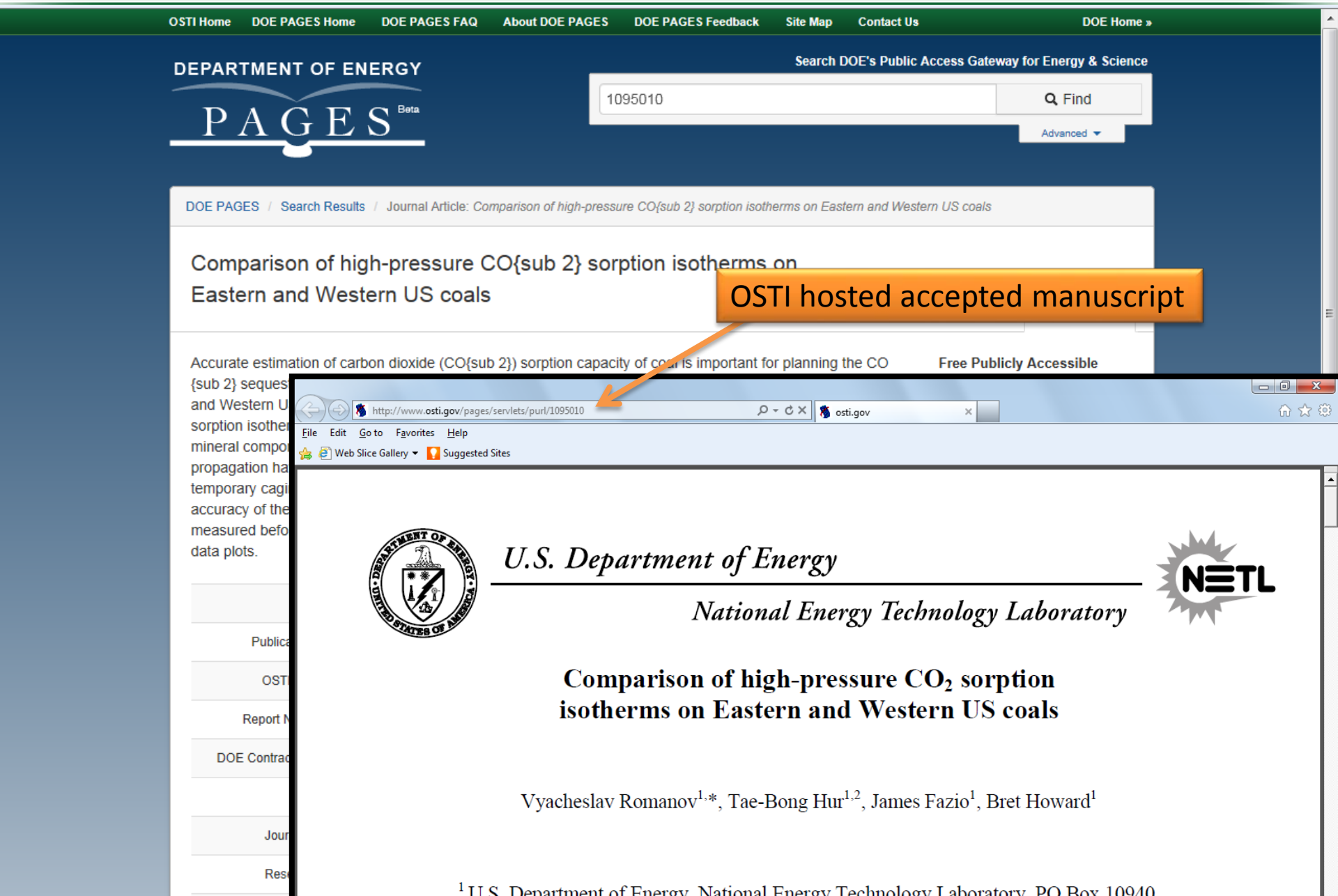
Chicago

Bibtex

EXPORT METADATA

EndNote

PAGES Item hosted by OSTI



PAGES Item with immediate public access

DEPARTMENT OF ENERGY

PAGES Beta

Search DOE's Public Access Gateway for Energy & Science

Start new search -- Place phrase in "double quotes" | Find | Advanced

Detection of Homologous Recombination Events in Bacterial Genomes

Citation Details

We study the detection of mutations, sequencing errors, and homologous recombination events (HREs) in a set of closely related microbial genomes. We base the model on single nucleotide polymorphisms (SNPs) and break the genomes into blocks to handle the rearrangement problem. Then we apply a dynamic programming algorithm to model whether changes within each block are likely a result of mutations, sequencing errors, or HREs. Results from simulation experiments show that we can detect 31%?61% of HREs and the precision of our detection is about 48%?90% depending on the rates of mutation and missing data. The HREfinder software for predicting HREs in a set of whole genomes is available as open source (http://sourceforge.net/projects/hrefinder/).

Authors:	Wang, Wei-Bung; Jiang, Tao; Gardner, Shea
Publication Date:	2013-10-07
OSTI Identifier:	OSTI ID: 1096427
DOE Contract Number:	AC52-07NA27344
Type:	Published Article
Journal Name:	PLoS ONE
Volume:	8
Issue:	10
Research Org:	Lawrence Livermore National Laboratory (LLNL), Livermore, CA (United States)

Free Publicly Accessible Full Text

Publisher's Version of Record at http://dx.doi.org/10.1371/journal.pone.0075230

Have feedback or suggestions for a way to improve these results? Let us know!

CITATION FORMATS

MLA
APA
Chicago
Bibtex

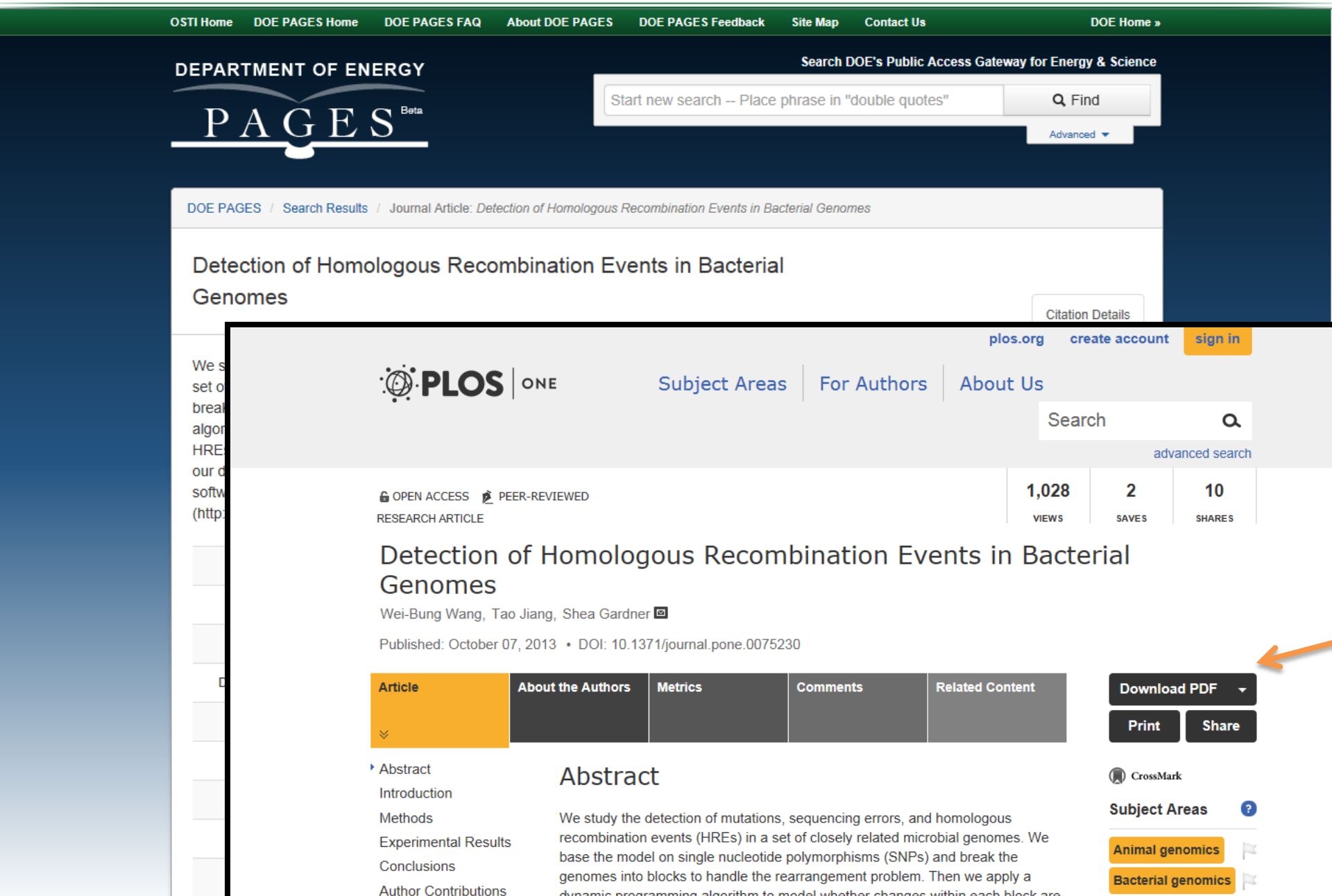
EXPORT METADATA

EndNote
Excel

SAVE / SHARE THIS RECORD

Send to Email

PAGES Item with immediate public access



CHORUS: Clearing House for Open Research of the U.S.*

- ~90 publishers are signatories to CHORUS, potentially covering ~90% of DOE peer-reviewed publications
- Progress to date: CHORUS is evolving; 7 publishers active in CHORUS pilot project, which contains ~4k journal records.
  - All publications in CHORUS will be available through PAGES
- Participating publishers agree to make articles resulting from federal funds publically available “after the determined embargo for each discipline and agency”
- CHORUS project provides freely consumable metadata for these articles, including seamless link to publisher VoR for tools like PAGES
- Articles resulting from federal funding are identified by additional **FundRef** metadata fields.
  - OSTI is a member of FundRef

* http://chorusaccess.org/about/

FundRef*

- Agreement by publishers to include two additional metadata fields for published articles
  1. funding organization
  2. grant/contract number (or equivalent)

- Authors are responsible for providing metadata to publishers at the time they submit article for consideration

- Articles from participating publishers that are tagged as having federal funding sources will be made available through CHORUS

* http://search.crossref.org/fundref

BACKUP

Participating CHORUS Publishers

1. Acoustical Society of America
2. Alliance of Crop, Soil, and Environmental Science Societies (ACSESS)
3. American Association for the Advancement of Science
4. American Association of Anatomists
5. American Association for Cancer Research
6. American Association of Physicists in Medicine
7. American Association of Physics Teachers
8. American Astronomical Society
9. American Chemical Society
10. American Crystallographic Association, Inc.
11. American College of Chest Physicians
12. American College of Physicians
13. American Dental Association
14. American Diabetes Association
15. American Geophysical Union
16. American Institute of Aeronautics and Astronautics
17. American Institute of Biological Sciences
18. American Institute of Physics
19. American Mathematical Society
20. American Meteorological Society
21. American Medical Association
22. American Nuclear Society
23. American Physical Society
24. American Physiological Society
25. American Psychiatric Publishing
26. American Psychological Association
27. American Society for Biochemistry and Molecular Biology
28. American Society of Agricultural & Biological Engineers
29. American Society of Agronomy
30. American Society of Civil Engineers
31. American Society of Mechanical Engineers
32. American Society for Microbiology
33. American Society of Plant Biologists
34. American Speech-Language-Hearing Association
35. Asociación Colombiana de Infectología
36. Association for Computing Machinery (ACM)
37. Association for Psychological Science
38. Association for Research in Vision and Ophthalmology
39. AVS: Science & Technology of Materials, Interfaces and Processing
40. Biophysical Society
41. Bioscientifica
42. Botanical Society of America
43. British Editorial Society of Bone & Joint Surgery
44. BMJ
45. Cambridge University Press
46. Columbia University Press
47. The Company of Biologists
48. Crop Science Society of America
49. Dove Medical Press
50. Duke University Press
51. Ecological Society of America
52. EDP Sciences
53. Elsevier
54. Emerald Group Publishing Limited
55. The Endocrine Society
56. Entomological Society of America
57. European Respiratory Society
58. Fabricators and Manufacturers Association, International
59. Genetics Society of America
60. GeoScienceWorld
61. Human Factors and Ergonomics Society
62. IEEE
63. iMedPub. Internet Medical Publishing
64. The Institute for Operations Research and the Management Sciences (INFORMS)
65. Institute of Physics Publishing
66. Institution of Engineering and Technology (IET)
67. Journal of Bone and Joint Surgery
68. Journal of Infection in Developing Countries
69. Journal of Rehabilitation Research and Development
70. Lynne Rienner Publishers, Inc.
71. Materials Research Society
72. McGraw-Hill
73. Mycological Society of America
74. New England Journal of Medicine
75. The Optical Society
76. Oxford University Press
77. The Physiological Society
78. Revista Medica de Risaralda
79. The Royal College of Psychiatrists
80. The Royal Society
81. Royal Society of Chemistry
82. Sage Publications
83. Society for the Advancement of Materials on Process Engineering
84. Society for the Study of Reproduction
85. Soil Science Society of America
86. Springer Science+Business Media LLC
87. Taylor & Francis
88. Thieme Publishers
89. University of Chicago Press
90. John Wiley & Sons
91. Wolters Kluwer Medical Research