CONNECTING SCIENTIFIC ARTICLES WITH RESEARCH DATA: NEW DIRECTIONS IN ONLINE SCHOLARLY PUBLISHING

IJsbrand Jan Aalbersberg[*]*, Judson Dunham, and Hylke Koers*

Elsevier, Radarweg 29, 1043 NX Amsterdam, The Netherlands
Email: IJ.J.Aalbersberg@elsevier,com

ABSTRACT

Researchers across disciplines are increasingly utilizing electronic tools to collect, analyze, and organize data. However, when it comes to publishing their work, there are no common, well-established standards on how to make that data available to other researchers. Consequently data is often not stored in a consistent manner, making it hard or impossible to find data sets associated with an article – even though such data might be essential to reproduce results or to perform further analysis. Data repositories can play an important role in improving this situation, offering increased visibility, domain-specific coordination, and expert knowledge on data management. As a leading STM publisher, Elsevier is actively pursuing opportunities to establish links between the online scholarly article and data repositories. This helps to increase usage and visibility for both articles and data sets, and also adds valuable context to the data. These data-linking efforts tie in with other initiatives at Elsevier to enhance the online article in order to connect with current researchers' workflows and to provide an optimal platform for the communication of science in the digital era.

Keywords: STM publishing, Data repositories, Data-linking, Innovation

1 INTRODUCTION

Driven by technological advancements enabling storage and sharing of large volumes of data, experimental data sets (possibly very large) have become an essential part of scientific research. Data in science is ubiquitous, with notable examples spanning a wide range of research disciplines from the Human Genome Project, to Earth Observations, to the Large Hadron Collider. Science, by its nature, produces a lot of data - and this is increasingly true now that many sensory data are born-digital and barriers to storing and processing this data are low.

Making research data available to other scholars provides an impetus to the advancement of science. First of all, it enables others to reproduce a scientific result to assure themselves of its validity – one of the cornerstones of the scientific method, yet often impossible to follow through when parts of the input data or computational methodology remain a black box. Secondly, data may be re-used for other purposes, sometimes unforeseen when the data was gathered. Re-use drives research efficiency by preventing duplication of work, but also opens the door to analyses that were not otherwise possible – in particular when different data sets are combined into a meta-analysis of sorts.

Despite the availability of basic enabling technologies, the potential for sharing research data is far from being fulfilled today. A recent study by PARSE.Insight [1] shows how researchers share their data in many ways: by email, on their university website, as supplementary material to a journal article, in an institutional repository, etc. This makes it hard for others to find data sets and, by the absence of clear and consistent metadata, to interpret them correctly. In addition, researchers are sometimes reluctant to share their data sets. There are several reasons for this: the additional work, worry about incorrect usage, the desire to "monetize" the value in data that was collected by hard work through a series of journal articles, or simply unawareness of existing possibilities.

What appears to be lacking to fully benefit from accessible research data is often organizational in nature: incentives to share data (in the form of academic credits, recognition, or otherwise), and common standards and processes that are widely accepted and consistently followed. Many stakeholders need to align to make that happen, and such a movement appears to be happening at present. An increasing number of funding agencies require their researchers to share data and/or have a data management plan. Domain-specific data repositories are taking up a role as centralized places to deposit and access research data. Organizations like DataCite [2] and the

ICSU World Data System [3] help to create overarching policies, views, or infrastructure to facilitate some of the work for individual data repositories, such as creating persistent identifiers, increasing discoverability, and establishing authority. Last, but not least, publishers are actively establishing connections between the scholarly record and data sets. This helps increase visibility and usage of data sets, integrates data sets into the existing researcher workflow, and provides accurate context to data sets – thereby addressing some of the issues that researchers face with sharing and re-using data. It is worth noting in this context that the international STM publishing community has issued a statement in 2007 to outline their view that "raw research data should be made freely available to all researchers" [4].

Scientific data repositories are numerous and diverse in character. A recent survey identified over a thousand scientific data repositories in Life Sciences alone [5]. The character of these repositories depends on the field and the intended audience - some data repositories just provide researchers a "safe harbor" to store their data sets for perpetuity, others actively curate the literature and organize data into authoritative information resources.

As a globally leading STM publisher, Elsevier has taken a prominent role in establishing (reciprocal) connections between the scholarly article and scientific data repositories. Such connections can take various forms, from clickable hyperlinks in the article text to interactive applications integrated into SciVerse® ScienceDirect® that pull data on-the-fly from a data resource on the web. What they have in common is the goal to present the reader with relevant, trustworthy data and information in the context of a research publication, to provide context to data sets, and to make it easer for researchers to find publications and data sets relevant to their work.

In the remainder of this article we will discuss data-linking initiatives at Elsevier. We will also describe the enabling technologies that Elsevier has invested in to allow for agile and collaborative development of data linking tools, and touch upon the vision underlying these efforts.

2 ELSEVIER'S ARTICLE OF THE FUTURE

The electronic age has brought profound changes to the way scientific research is conducted and captured. Researchers increasingly use electronic tools to perform measurements, and to analyze, organize, and share their material. Consequently research output is increasingly diverse and "rich" in an electronic sense – including data sets, video and other multimedia files, computer code, etc. However, when it comes to publishing, the scientific article – as a vehicle to communicate that research - has shown little adaptation and a scientist often finds herself reducing valuable scientific output to "ink on paper" – which the reader then has to reconstruct to take full advantage of the insights the author wanted to share (many a researcher will recognize the frustration of having to re-key a data table, or use a ruler to determine the location of a peak on a data plot.)

Improvements to the scholarly article over the last few decades have been mostly in terms of delivery (electronically), discoverability (full-text search), as well as a number of smaller-scale, specific enhancements such as the possibility to upload supplementary data. However, in terms of structure and shape, the current article is by and large the same as in the first scholarly journals of the 17th century. In order to address this growing mismatch between the frozen "article" concept, and the evolving workflows and needs of researchers, Elsevier has initiated the "Article of the Future" – a project to rethink the scientific article in the electronic age.

The Article of the Future project aims to offer an optimal platform to communicate science in today's digital world. From this starting point, the concept has been developed in close collaboration with the scientific community, involving feedback from several hundreds of researchers. Very early on in the development, it was recognized that the greatest additional value lies in domain-specific enhancements. Different scientific communities use different tools, are used to different data standards, and might have different attitudes towards communicating research output – so one clearly has to go beyond "one size fits all" to really connect with, and support, the workflow of individual researchers. The Article of the Future improves the online article in essentially three directions:

- Presentation – offering an optimal online browsing and reading experience, which is a basic requirement for online reading and for any further enhancements.
- Content – supporting a richer pallet of author-delivered material, including multimedia files, scientific data and computer code.

- Context – connecting the online article to trustworthy scientific resources to present the reader with relevant information in the context of the article.

A perhaps naïve metaphor for the enhanced contextualization of articles is to think about the article as a roundabout rather than the traditional one-way street. Establishing connections between the article and data repositories perfectly fits into this vision, making the Article of the Future a natural format for seamless integration of data-linking tools.

Initially introduced for Cell Press journals in 2009, the Article of the Future concept was expanded to other disciplines from 2010 onwards. After an initial series of prototype articles on www.articleofthefuture.com, the first phase of this concept has been implemented across Elsevier's publication platform SciVerse ScienceDirect at the time of writing (January 2012). Further enhancements are expected later in 2012.

3 LINKING SCIVERSE SCIENCEDIRECT AND DATA REPOSITORIES

SciVerse ScienceDirect supports a number of methods to link with data repositories, including author-based tagging schemes, automatically generated data-linking banners (see also [6]) and data-linking applications. The most appropriate choice depends on the nature of the data and the database, and on how the information is best presented to the reader.

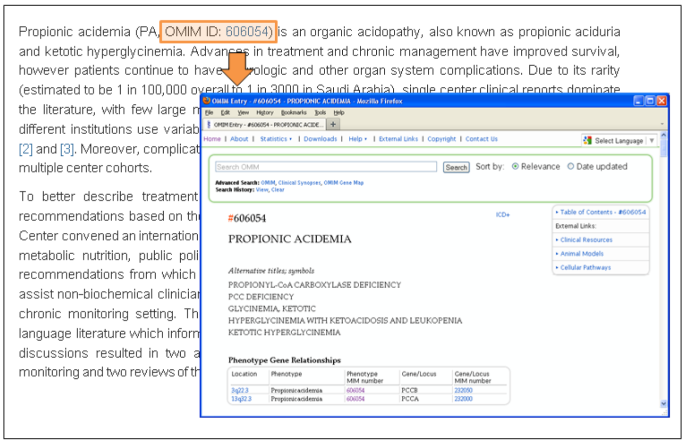
The first method that Elsevier employs to link out to a data repository is by asking authors to explicitly tag entities for which data is available at a repository, for example "OMIM ID: 606054" to refer to a specific data record at the Online Mendelian Inheritance in Man (OMIM) data repository. This tag will be recognized during the publication production process, and show up as a hyperlink in the online article, pointing to the relevant data record.

The second possibility relies on an automatic, on-the-fly linking banner service that has been developed for this purpose. Here a selected database is queried when a reader opens up an online article on SciVerse ScienceDirect. If the data repository recognizes the article DOI, a banner image is returned to indicate that the repository holds relevant data records for this article. Clicking on the banner then directs the reader to those records. A key benefit of this scheme is that it allows for data to be connected to the article after publication (possibly even years thereafter). This is particularly useful to help make past data available using current technology, and also accommodates authors who prefer to make their data available only after an initial embargo or delay.

A third, broad category of data-linking methods lies in utilizing Elsevier's SciVerse Application Framework [7] to create dedicated applications that connect with online data repositories and display relevant data and information alongside the online article for interactive exploration. This framework – which will be discussed in detail in the next section - opens the door to the wide range of possibilities for interactive exploration of data in the context of the online article.

At the time of writing, SciVerse ScienceDirect features over 20 linking schemes and data-linking applications to connect articles to scientific data repositories. Covering the wide gamut of subject areas, linked data repositories include the Protein Data Bank [8], Encyclopedia of Life [9], Cambridge Crystallographic Data Center [10], EarthChem [11], PANGAEA [12], the SIMBAD Astronomical Database [13], ClinicalTrials.gov, Data.gov – and many others.[1]

[1] For an up-to-date overview of author-tagged linking schemes and SciVerse (data-linking) applications, please visit http://www.elsevier.com/databaselinking and http://www.applications.sciverse.com/action/gallery, respectively.



Propionic acidemia (PA, OMIM ID: 606054) is an organic acidopathy, also known as propionic aciduria
and ketotic hyperglycinemia. Advances in treatment and chronic management have improved survival,
however patients continue to have ologic and other organ system complications. Due to its rarity
(estimated to be 1 in 100,000 overall to 1 in 3000 in Saudi Arabia), single center clinical reports dominate
the literature, with few large m
different institutions use variab
[2] and [3]. Moreover, complicat
multiple center cohorts.

To better describe treatment
recommendations based on th
Center convened an internation
metabolic nutrition, public poli
recommendations from which
assist non-biochemical clinician
chronic monitoring setting. Th
language literature which inform
discussions resulted in two a
monitoring and two reviews of th

OMIM Entry - #606054 - PROPIONIC ACIDEMIA - Mozilla Firefox
File Edit View History Bookmarks Tools Help
OMIM Entry - #606054 - PROPIONIC ACIDE...
Home | About | Statistics | Downloads | Help | External Links | Copyright | Contact Us
Select Language
Search OMIM
Search
Sort by: ● Relevance ○ Date updated
Advanced Search: OMIM, Clinical Synopses, OMIM Gene Map
Search History: View, Clear
ICD+
Table of Contents - #606054
External Links:
Clinical Resources
Animal Models
Cellular Pathways

#606054

PROPIONIC ACIDEMIA

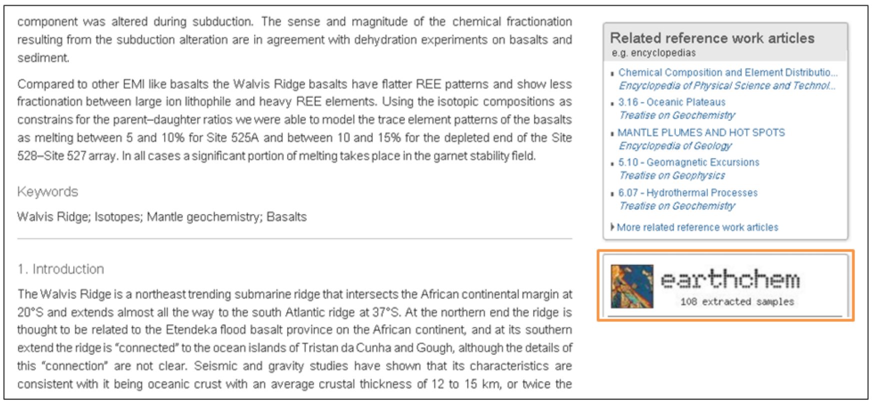
Alternative titles; symbols

PROPIONYL-CoA CARBOXYLASE DEFICIENCY
PCC DEFICIENCY
GLYCINEMIA, KETOTIC
HYPERGLYCINEMIA WITH KETOACIDOSIS AND LEUKOPENIA
KETOTIC HYPERGLYCINEMIA

Phenotype Gene Relationships

Location	Phenotype	Phenotype MIM number	Gene/Locus	Gene/Locus MIM number
3q22.3	Propionicacidemia	606054	PCCB	232050
13q32.3	Propionicacidemia	606054	PCCA	232000

Figure 1. Screenshot of an online article page on SciVerse ScienceDirect showing an author-tagged link to the OMIM data repository (see http://dx.doi.org/10.1016/j.ymgme.2011.08.007). The inset shows the landing page for the data record at OMIM. Here, as in the other figures, the orange boxes and arrows (pointing from a link to its target) are for illustration purposes only; these are not displayed on SciVerse ScienceDirect.



component was altered during subduction. The sense and magnitude of the chemical fractionation resulting from the subduction alteration are in agreement with dehydration experiments on basalts and sediment.

Compared to other EMI like basalts the Walvis Ridge basalts have flatter REE patterns and show less fractionation between large ion lithophile and heavy REE elements. Using the isotopic compositions as constrains for the parent–daughter ratios we were able to model the trace element patterns of the basalts as melting between 5 and 10% for Site 525A and between 10 and 15% for the depleted end of the Site 528–Site 527 array. In all cases a significant portion of melting takes place in the garnet stability field.

Keywords

Walvis Ridge; Isotopes; Mantle geochemistry; Basalts

1. Introduction

The Walvis Ridge is a northeast trending submarine ridge that intersects the African continental margin at 20°S and extends almost all the way to the south Atlantic ridge at 37°S. At the northern end the ridge is thought to be related to the Etendeka flood basalt province on the African continent, and at its southern extend the ridge is “connected” to the ocean islands of Tristan da Cunha and Gough, although the details of this “connection” are not clear. Seismic and gravity studies have shown that its characteristics are consistent with it being oceanic crust with an average crustal thickness of 12 to 15 km, or twice the

Related reference work articles
e.g. encyclopedias

- Chemical Composition and Element Distributio... *Encyclopedia of Physical Science and Technol...*
- 3.16 - Oceanic Plateaus *Treatise on Geochemistry*
- MANTLE PLUMES AND HOT SPOTS *Encyclopedia of Geology*
- 5.10 - Geomagnetic Excursions *Treatise on Geophysics*
- 6.07 - Hydrothermal Processes *Treatise on Geochemistry*

More related reference work articles

Figure 2. Screenshot of an EarthChem linking banner as displayed next to an online article on SciVerse ScienceDirect (see http://dx.doi.org/10.1016/j.chemgeo.2010.02.010).

4 THE SCIVERSE APPLICATION FRAMEWORK AND DATA-LINKING APPLICATIONS

4.1 The SciVerse Application Framework as an Enabling Technology

The SciVerse Application Framework is a collection of technologies that, amongst others, enable dynamic integration of data and software from third-party sources into Elsevier’s publishing and search platforms SciVerse ScienceDirect, SciVerse Scopus®, and SciVerse Hub. The component systems to support these integrations provide services ranging from platform integration, search and retrieval APIs [2] and content syndication, to advanced services like entity extraction and text mining.

The core component of the SciVerse Application Framework is the platform integration infrastructure, or Framework API, which allows for third-party data and tools to integrate dynamically into the SciVerse user interface. To accomplish this, SciVerse uses a standards-based, open source gadget framework: Apache Shindig, an OpenSocial container which allows for applications, or “gadgets”, to be integrated into container web applications. Elsevier has extended its implementation of this software in a few key ways to offer a range of functionality especially suited for linking with data sets:

[2] APIs, short for Application Programming Interfaces, are machine-readable interfaces that allow for programmatic access to content and services.

- Platform and content integration – applications running within SciVerse ScienceDirect have access to the full-text scientific article, the document metadata, and user input such as search terms.
- User interface extensions – applications have access to a wide array of functions to enable rich user experiences and integration into the surrounding user interface, such as the ability to create links within documents and load applications dynamically within the page.
- Interaction with external resources – applications can access information resources on the web to combine that with Elsevier content into a form most useful to readers.

A cornerstone of the SciVerse Application Framework is the set of SciVerse Content APIs, which allow programmatic access to a broad range of scholarly content and bibliometrics within SciVerse ScienceDirect and SciVerse Scopus. The APIs allow for real time query and retrieval of search results, abstracts, full-text articles, document metadata, author and affiliation data and citation metrics. Created for easy re-use to facilitate new application development, these tools are now frequently used in text analysis and matching operations to identify and link data references within articles.

In cases where on-the-fly analysis tools are too slow, and pre-processing is required, Elsevier also provides a Content Syndication service to deliver large amounts of full-text articles and books to data linking partners and application developers. This service allows for rapid delivery of configurable collections of full-text content in bulk, either downloadable via FTP or delivered on hard disk. New content can be updated via FTP, with configurable delivery schedules that can provide regular content updates as often as hourly, to enable partners to access newly published full-text content with minimal delays.

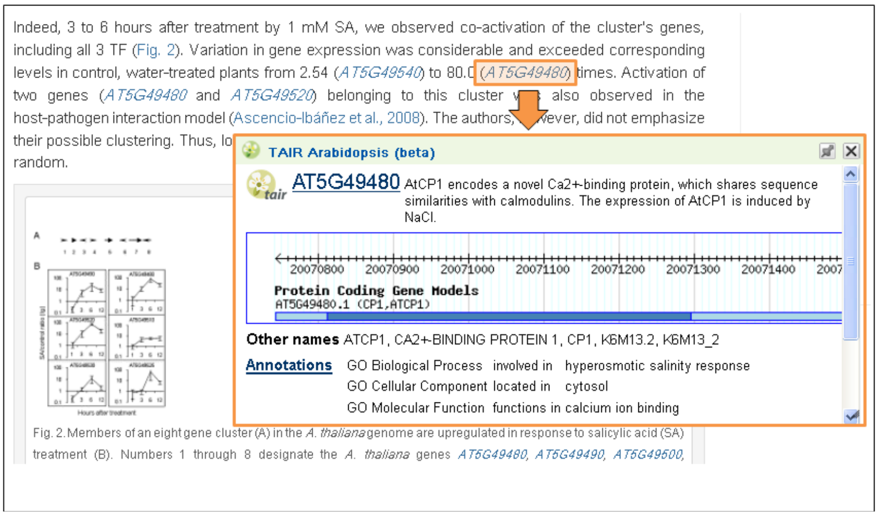
Beyond these basic components, Elsevier is also expanding its services into more advanced areas by creating "Developer Services". These are meant to enable rapid development and deployment of new data-linking opportunities as they arise. These building blocks also help to let potential partners focus on the scientific implementation of a data application, rather than on the technical details. As an example of this, a prototype entity extraction service was recently made available. This service can be configured with a lexicon containing phrases associated with the entities to be identified. It can then be passed an arbitrary chunk of text in which it identifies entities from the associated lexicon. The response from this service contains the list of entities matched in the text, the number of occurrences, and the location of those occurrences. In this way, data linking applications can be created with very little effort, often requiring little more than the transfer of a lexicon of unique terms or identifiers as a basis.

4.2 Some Examples of Data-Linking Applications on SciVerse ScienceDirect

LIPID Structures & TAIR Arabidopsis

Elsevier recently collaborated with the community-organized LIPID MAPS [14] and TAIR [15] (The Arabidopsis Information Resource) data repositories to develop two applications for researchers in biology. Both applications operate in a similar manner: they identify terms within articles on SciVerse ScienceDirect, pull key information into the article, and link to the associated data record pages – thereby helping researchers to quickly access valuable reference information on Arabidopsis loci (TAIR) or lipids (LIPID Structures).

These applications provide examples of the collaborative development discussed in the previous section, allowing for rapid development and deployment. They use a centralized, configurable text-mining service created by Elsevier. This enables data repositories to focus on their key strengths such as localized collection, aggregation, enhancement and storage of domain-specific data, and selection of the most relevant information for online readers.



Indeed, 3 to 6 hours after treatment by 1 mM SA, we observed co-activation of the cluster's genes, including all 3 TF (Fig. 2). Variation in gene expression was considerable and exceeded corresponding levels in control, water-treated plants from 2.54 (*AT5G49540*) to 80.0 (*AT5G49480*) times. Activation of two genes (*AT5G49480* and *AT5G49520*) belonging to this cluster w[...] also observed in the host-pathogen interaction model (Ascencio-Ibáñez et al., 2008). The authors, [...]ever, did not emphasize their possible clustering. Thus, lo[...] random.

TAIR Arabidopsis (beta)

AT5G49480 AtCP1 encodes a novel Ca2+-binding protein, which shares sequence similarities with calmodulins. The expression of AtCP1 is induced by NaCl.

Protein Coding Gene Models
AT5G49480.1 (CP1,ATCP1)

Other names ATCP1, CA2+-BINDING PROTEIN 1, CP1, K6M13.2, K6M13_2

Annotations

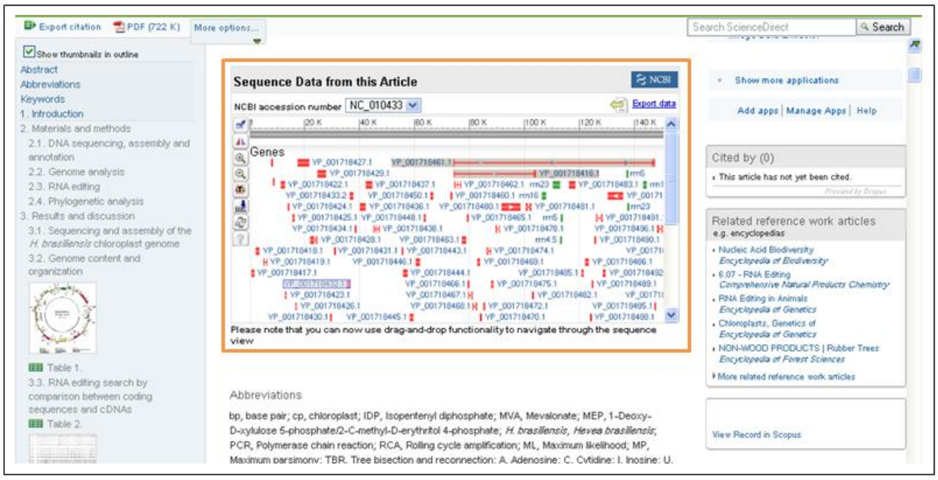
GO Biological Process	involved in	hyperosmotic salinity response
GO Cellular Component	located in	cytosol
GO Molecular Function	functions in	calcium ion binding

Fig. 2. Members of an eight gene cluster (A) in the *A. thaliana* genome are upregulated in response to salicylic acid (SA) treatment (B). Numbers 1 through 8 designate the *A. thaliana* genes *AT5G49480*, *AT5G49490*, *AT5G49500*,

Figure 3. Screenshot showing a locus recognized by the TAIR application, and the information panel after being opened by the reader (see http://dx.doi.org/10.1016/j.gene.2011.09.023).

The Genome Viewer

The Genome Viewer, developed in collaboration with NCBI, combines author-tagged entities with SciVerse's capabilities to interact with external resources and to add a layer of interactivity that lets readers interactively explore data rather than having to digest a long list with information. The application recognizes NCBI Accession Numbers for genetic sequences, collects sequence data from GenBank [16], and collects this in an interactive sequence viewer. Users can easily find locations of specific interest, change the visualization, or download sequence data from within the application.



Export citation PDF (722 K) More options...

Search ScienceDirect Search

☑ Show thumbnails in outline

Abstract
Abbreviations
Keywords
1. Introduction
2. Materials and methods
2.1. DNA sequencing, assembly and annotation
2.2. Genome analysis
2.3. RNA editing
2.4. Phylogenetic analysis
3. Results and discussion
3.1. Sequencing and assembly of the *H. brasiliensis* chloroplast genome
3.2. Genome content and organization
Table 1.
3.3. RNA editing search by comparison between coding sequences and cDNAs
Table 2.

Sequence Data from this Article

NCBI accession number NC_010433

Export data

Please note that you can now use drag-and-drop functionality to navigate through the sequence view

Abbreviations

bp, base pair; cp, chloroplast; IDP, Isopentenyl diphosphate; MVA, Mevalonate; MEP, 1-Deoxy-D-xylulose 5-phosphate/2-C-methyl-D-erythritol 4-phosphate; *H. brasiliensis*, *Hevea brasiliensis*; PCR, Polymerase chain reaction; RCA, Rolling cycle amplification; ML, Maximum likelihood; MP, Maximum parsimony; TBR, Tree bisection and reconnection; A, Adenosine; C, Cytidine; I, Inosine; U,

Show more applications

Add apps | Manage Apps | Help

Cited by (0)

- This article has not yet been cited.

Related reference work articles
e.g. encyclopedias

- Nucleic Acid Biodiversity *Encyclopedia of Biodiversity*
- 6.07 - RNA Editing *Comprehensive Natural Products Chemistry*
- RNA Editing in Animals *Encyclopedia of Genetics*
- Chloroplasts, Genetics of *Encyclopedia of Genetics*
- NON-WOOD PRODUCTS | Rubber Trees *Encyclopedia of Forest Sciences*

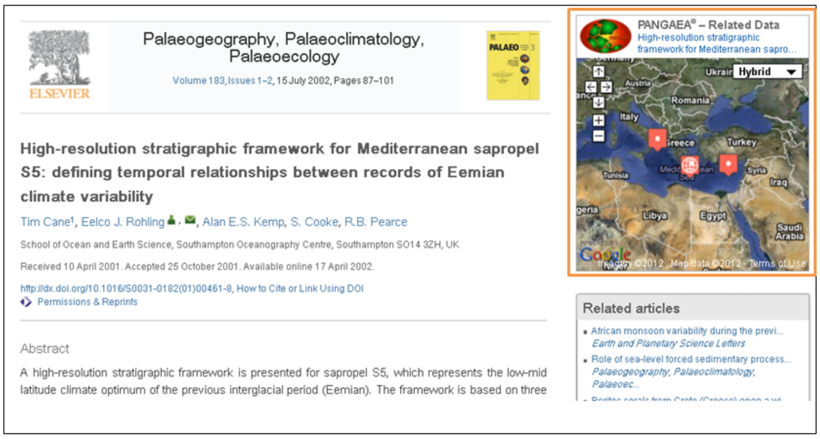
More related reference work articles

View Record in Scopus

Figure 4. Screenshot showing genetic sequence data visualized by the Genome Viewer (see http://dx.doi.org/10.1016/j.gene.2011.01.002).

PANGAEA

Elsevier and PANGAEA have built an advanced linking service between SciVerse ScienceDirect and the PANGAEA data repository for Earth Sciences research. Authors who submit a paper to a participating journal are encouraged to submit their raw data sets to PANGAEA, where they are archived and assigned a unique, persistent identifier. When the paper is published online, the reader will see an interactive map application that visualizes the geographical locations of the data sets at PANGAEA and offers links to the data records.



Palaeogeography, Palaeoclimatology, Palaeoecology

Volume 183, Issues 1–2, 15 July 2002, Pages 87–101

High-resolution stratigraphic framework for Mediterranean sapropel S5: defining temporal relationships between records of Eemian climate variability

Tim Cane[1], Eelco J. Rohling, Alan E.S. Kemp, S. Cooke, R.B. Pearce

School of Ocean and Earth Science, Southampton Oceanography Centre, Southampton SO14 3ZH, UK

Received 10 April 2001. Accepted 25 October 2001. Available online 17 April 2002.

http://dx.doi.org/10.1016/S0031-0182(01)00461-8, How to Cite or Link Using DOI

Permissions & Reprints

Abstract

A high-resolution stratigraphic framework is presented for sapropel S5, which represents the low-mid latitude climate optimum of the previous interglacial period (Eemian). The framework is based on three

PANGAEA® – Related Data

High-resolution stratigraphic framework for Mediterranean sapro...

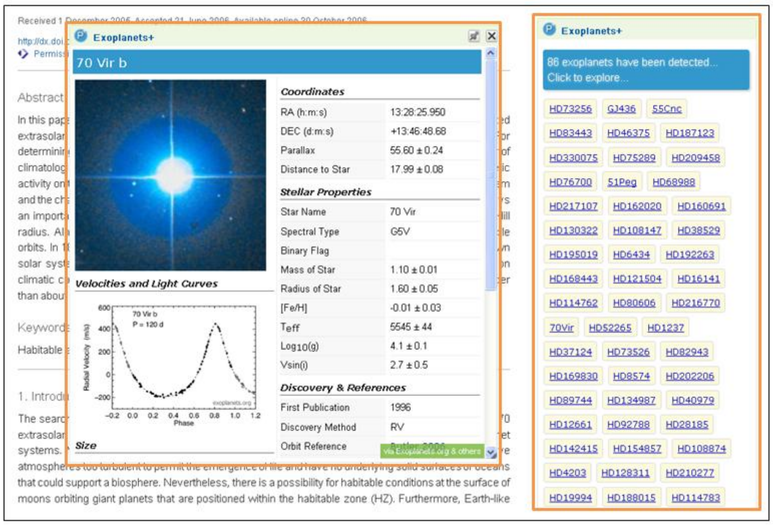
Related articles

- African monsoon variability during the previ... *Earth and Planetary Science Letters*
- Role of sea-level forced sedimentary process... *Palaeogeography, Palaeoclimatology, Palaeoec...*

Figure 5. Screenshot showing the PANGAEA map viewer with the geographical locations of data records at PANGAEA (see http://dx.doi.org/10.1016/S0031-0182(01)00461-8).

Exoplanets+

The Exoplanets+ Application was one of the contestants for Elsevier's "Apps for Science" contest. The application searches through full-text articles for exoplanet (extra-solar planet) names that it recognizes from several astronomical data repositories including Exoplanets.org and SIMBAD [13]. If exoplanets are found, the application opens up in the right-hand pane to alert the reader that additional information is available. Clicking on an exoplanet name opens a panel with a compilation of key information, and links to the underlying data repositories.



Exoplanets+

70 Vir b

Coordinates	
RA (h:m:s)	13:28:25.950
DEC (d:m:s)	+13:46:48.68
Parallax	55.60 ± 0.24
Distance to Star	17.99 ± 0.08

Stellar Properties	
Star Name	70 Vir
Spectral Type	G5V
Binary Flag	
Mass of Star	1.10 ± 0.01
Radius of Star	1.60 ± 0.05
[Fe/H]	-0.01 ± 0.03
Teff	5545 ± 44
Log10(g)	4.1 ± 0.1
Vsin(i)	2.7 ± 0.5

Discovery & References	
First Publication	1996
Discovery Method	RV
Orbit Reference	Butler 2006

via Exoplanets.org & others

that could support a biosphere. Nevertheless, there is a possibility for habitable conditions at the surface of moons orbiting giant planets that are positioned within the habitable zone (HZ). Furthermore, Earth-like

Exoplanets+

86 exoplanets have been detected... Click to explore...

HD73256 GJ436 55Cnc
HD83443 HD46375 HD187123
HD330075 HD75289 HD209458
HD76700 51Peg HD68988
HD217107 HD162020 HD160691
HD130322 HD108147 HD38529
HD195019 HD6434 HD192263
HD168443 HD121504 HD16141
HD114762 HD80606 HD216770
70Vir HD52265 HD1237
HD37124 HD73526 HD82943
HD169830 HD8574 HD202206
HD89744 HD134987 HD40979
HD12661 HD92788 HD28185
HD142415 HD154857 HD108874
HD4203 HD128311 HD210277
HD19994 HD188015 HD114783

Figure 6. Screenshot showing a list of exoplanet names recognized by the Exoplanet+ application, and the information panel after being opened by the reader (see http://dx.doi.org/10.1016/j.pss.2006.06.022).

5 CONCLUSION

In a time where scientists are increasingly utilizing electronic tools for their research tasks - think only of the extensive use of automated data collection and processing pipelines, or the ubiquity of software tools to analyze and share research material - new or improved ways to disseminate scientific output are emerging. In particular, several stakeholders are actively encouraging the sharing of research data through scientific data repositories –

so as to benefit from increased visibility, domain-specific coordination, and expert knowledge on data management.

Recognizing the need to support changing workflows and researcher needs (in both the author and reader role) Elsevier has taken several initiatives to bring additional value to its online articles on SciVerse ScienceDirect. These include the Article of the Future project, focusing on improved presentation, content and context, and the SciVerse Application Framework, which enables Elsevier and partner organizations to develop specific, interactive tools that the reader can access from within the context of the online article.

Building on these foundations, Elsevier is actively establishing connections between online articles and scientific data repositories. This improves the user experience for readers of SciVerse ScienceDirect by providing simple, one-click access to trustworthy and relevant data. At the same time, this program improves visibility and usage of data repositories, and places data sets in perspective: journal articles often contain essential information about data sets – how were they accumulated, what are their limitations, which conclusions have been drawn from them, etc. – that is essential for correct interpretation and consistent re-use. In this manner, connections between the scientific articles and data repositories add value on both sides, and contribute to a better infrastructure for the dissemination of science in the electronic age.

Elsevier is keen on further expanding its range of data-linking schemes and applications, and welcomes collaboration with interested parties.

6 REFERENCES

[1] http://www.parse-insight.eu/
[2] http://www.datacite.org
[3] http://www.icsu-wds.org/
[4] http://www.stm-assoc.org/2007_11_01_Brussels_Declaration.pdf
[5] Laura Haak Marcial, Bradley M. Hemminger (2010): Scientific Data Repositories on the Web: An Initial Survey. JASIST 61 (10) 2029-2048. doi:10.1002/asi.21339
[6] IJsbrand Jan Aalbersberg, Ove Kähler (2011): Supporting Science through the Interoperability of Data and Articles. D-Lib Magazine 17 (1/2). doi:10.1045/january2011-aalbersberg
[7] http://www.applications.sciverse.com
[8] http://www.rcsb.org
[9] http://www.eol.org
[10] http://www.ccdc.cam.ac.uk
[11] http://www.earthchem.org
[12] http://www.pangaea.de
[13] http://simbad.u-strasbg.fr/simbad
[14] http://www.lipidmaps.org
[15] http://www.arabidopsis.org
[16] http://www.ncbi.nlm.nih.gov/genbank