# Deep Indexing: Harnessing the Power of Data Discovery

Mark Hyer VP, Secondary Publishing

ProQuest®

Start here.

# Agenda

- Attempts to Search Tables and Figures Today
- Deep Indexing- A Quick Overview
- Deep Indexing Projects
- Questions for the Group

# Search and Retrieval

- Explosion of content available through:
  - Abstract and Indexing
  - Full Text
  - Dissertations
  - **Government Documents**
    - **Ex. Insect Resistant Corn Stats in South Dakota**
  - Special Collections
  - Social Networks
  - Special Collections
  - Open Web……

Insect resistant corn south dakota - Google Search - Windows Internet Explorer

http://www.google.com/search?hl=en&rls=com.microsoft%3Aen-us%3AIE-SearchBox&q=Insect+resistant+corn+south+dakota

Insect resistant corn south dakota - Google Search

Web Images Maps News Shopping Gmail more ▼

Google

Insect resistant corn south dakota

Search

Advanced Search
Preferences

Web

Results **1** - **10** of about **453,000** for Ins

Bio-9--**Insect-resistant** Crops Through Genetic Engineering
Genetically engineered **insect-resistant corn** is providing a new defense against ..... Missouri, Nebraska, North **Dakota**, Ohio, **South Dakota**, and Wisconsin. ...
www.biotech.iastate.edu/biotech_info_series/bio9.html - 33k - Cached - Similar pages

[PDF] Bt **Corn** and **Insect Resistance** Management - What Are They? (SSMG-33)
File Format: PDF/Adobe Acrobat - View as HTML
Growers must practice **insect resistance**. management (IRM) to ensure that this technology will be .... **Dakota Corn** Utilization Council, **South Dakota** Soybean ...
www.ppi-far.org/.../b369c6dbe705dd13852568e3000de93d/62f1f3f2f34e04e2852569db0032bbde/$FILE/SSMG-33.pdf - Similar pages

AgBioForum 10(1): **Insect Resistance** Management Plans: The Farmers ...
Key words: Bt **corn**, **insect resistance** management, refuge. ... the planting practices of **corn**-soybean producers in **South Dakota** for Bt **corn resistant** to ECB. ...
www.agbioforum.org/v10n1/v10n1a04-alexander.htm - 64k - Cached - Similar pages

Bacillus thuringiensis aizawai,PS811 Cry1F and Cry35Ab1 proteins ...
... Mycogen Brand B.t. moCry1F **Insect Resistant Corn** Seed (EPA File Symbol 68467-U), ... Minnesota, Mississippi, Nebraska, North **Dakota**, Ohio, **South Dakota**, ...
www.epa.gov/opp00001/biopesticides/ingredients/fr_notices/frnotices_006491.htm - 20k - Cached - Similar pages

Transgenic **insect-resistant corn** affects the fourth trophic level ...
Transgenic **insect-resistant corn** affects the fourth trophic level: ..... Kfir R, Gouws J, Moore **SD** (1993) Biology of Tetrastichus ...
www.springerlink.com/index/5Y90QGRCK8RWT6AY.pdf - Similar pages

**CORN** Newsletter: 2008-19, June 23, 2008 - July 1, 2008

**South**
We offer
Find **Sou**
**SouthDa**

Insects
ClipArt -
Insects
www.live.

**Dakota**
Bargain
Save on
BizRate.

Insect resistant corn south dakota % - Google Scholar - Windows Internet Explorer

http://scholar.google.com/scholar?hl=en&lr=&scoring=r&q=Insect+resistant+corn+south+dakota+%25&as_ylo=2003

Insect resistant corn south dakota % - Google Scholar

Google Scholar BETA

Web Images Video News Maps more »

Insect resistant corn south dakota % | since 2003 | Search

Advanced Scholar Search
Scholar Preferences
Scholar Help

**Scholar** All articles - **Recent articles** Results **1** - **10** of about **4,490** for Insec

[CITATION] Validation of a novel resistance monitoring technique for **corn** rootworm (Coleoptera: Chrysomelidae) ...
TM Nowatzki, SA Lefko, RR Binning, SD Thompson, TA ... - Journal of Applied Entomology, 2008 - Blackwell Synergy
... Lefko SA, Nowatzki TM, Thompson **SD**, Binning RR ... Field measures of western **corn** rootworm (Coleoptera ... **Insect resistance** to transgenic Bt crops: lessons from the ...
Cited by 1 - Related Articles - Web Search - BL Direct

[CITATION] Characterizing laboratory colonies of western **corn** rootworm (Coleoptera: Chrysomelidae) selected for ...
SA Lefko, TM Nowatzki, SD Thompson, RR Binning, MA ... - Journal of Applied Entomology, 2008 - Blackwell Synergy
... TJ, Lefko SA, Thompson **SD**, Binning RA ... explicit model simulating western **corn** rootworm (Coleoptera: Chrysomelidae) adaptation to **insect-resistant** maize ...
Cited by 1 - Related Articles - Web Search - BL Direct

Weed Population Dynamics after Six Years under Glyphosate-and Conventional Herbicide-based Weed ...
P Westra, RG Wilson, SD Miller, PW Stahlman, GW ... - Crop Science, 2008 - Crop Sci Soc America
... was the dominant genetic trait, followed by **insect resistance**. ... 0.90 ha, using exclusively glyphosate-**resistant** crops in ... plots were either continuous **corn** or a ...
Web Search

Arthropod pest management in organic crops - all 4 versions »
G Zehnder, GM Gurr, S Kühne, MR Wade, SD Wratten, ... - Annual Review of Entomology, 2007 - Annual Reviews
... Recent studies have shown that plant **resistance** to **insect** and disease ... source and soil management history effects influencing the European **corn** borer (ECB ...
Cited by 14 - Related Articles - Web Search - BL Direct

Prairie Grasses as Hosts of the Northern **Corn** Rootworm (Coleoptera: Chrysomelidae) - all 4 versions »
IO Oyediran, BW French, TL Clark, KE Dashiell, BE ... - Environmental Entomology, 2008 - bioone.org
... Brooking, **SD**. ... Storer NP A spatially explicit model simulating western **corn** rootworm (Coleoptera: Chrysomelidae) adaptation to **insect-resistant** maize. ...

Biotech :: General Information - Windows Internet Explorer

http://www.sdcorn.org/biotech/

Biotech :: General Information

**Source:** USDA National Agricultural Statistics Service, Acreage Report, June 30, 2003

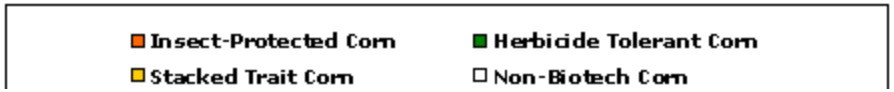
In South Dakota, however, 75% of all corn produced is biotech corn. That's a greater percentage than any other state.

### US Biotech Corn Production By State

| State | 2000 | 2001 | 2002 | 2003 |
|---|---|---|---|---|
| Illinois | 17% | 16% | 22% | 28% |
| Indiana | 11% | 12% | 13% | 16% |
| Iowa | 30% | 32% | 41% | 45% |
| Kansas | 33% | 38% | 43% | 47% |
| Michigan | 12% | 17% | 22% | 35% |
| Minnesota | 37% | 36% | 44% | 53% |
| Missouri | 28% | 32% | 34% | 42% |
| Nebraska | 34% | 34% | 46% | 52% |
| Ohio | 9% | 11% | 9% | 9% |
| South Dakota | 48% | 47% | 66% | 75% |
| Wisconsin | 18% | 18% | 26% | 32% |
| Other | 17% | 20% | 27% | 36% |
| **Total US** | **25%** | **26%** | **34%** | **40%** |

**Source:** USDA National Agricultural Statistics Service, Acreage Report, June 30, 2003

South Dakota has readily adopted the production of biotech corn. While the production of insect-protected corn has stayed about the same, South Dakota has increased the amount of herbicide tolerant corn and stacked trait corn:



National Agricultural Library: NAL Catalog (AGRICOLA) - Windows Internet Explorer

http://agricola.nal.usda.gov/cgi-bin/Pwebrecon.cgi?Search_Arg=insect+resistant+corn+south+dakota&DB=local&CNT=25&Search_Code=GKEY&STARTDB=AGRIDB

National Agricultural Library: NAL Catalog (AGRICOLA)

USDA United States Department of Agriculture
National Agricultural Library

Catalog Home | About the Catalog | New Search | Databases | History | Patron Info | Login | How to Request | Help | Con

**Database Name:** Article Citation Database

**Your search resulted in no hits.For assistance, click the Help or Contact Us buttons in the top toolbar.**

Keyword Search | **Basic Search** | Advanced Search

**Find:** insect resistant corn south dakota

**Basic Limit:** None

**In:**
- Title Begins with...
- Subject Browse
- Author Browse
- Journal Title Begins with...
- SuDoc Number
- Name/Title Browse

Search

25 records per page

Clear

**Basic Search Hints**

**General**

- If your search fails, check your spelling.

Science.gov : USA.gov for Science - Government Science Portal - Windows Internet Explorer

http://www.science.gov/

Science.gov : USA.gov for Science - Government Sci...

# science.gov

USA.gov for SCIENCE

Version 4.0

Site Map ★ Index ★ Alerts ★ Help ★ Contact Us ★ About Science.gov ★ Communications ★ Alliance Only

## Science.gov Co-Chair Named FLICC Federal Librarian of the Year

Science.gov is a gateway to over 50 million pages of authoritative selected science information provided by U.S. government agencies, including research and development results

**Featured Search**

salmonella

• Archive •

insect resistant corn

Begin Search

Use quotation marks for all phrases

More Search Tips

Advanced Search

## Explore Selected Science Web Sites *by Topic*

**Agriculture & Food**
Food Safety, Gardening, Pesticides, Veterinary Science ...

**Applied Science & Technologies**
Biotechnology, Electronics, Engineering, Transport ...

**Astronomy & Space**
Exploration, Planets, Space Technologies ...

**Biology & Nature**
Animals & Plants, Ecology, Genetics, Pest Control ...

**Computers & Communication**
Networks, Hardware, Software ...

**Earth & Ocean Sciences**
Land, Maps, Natural Disasters, Oceans, Weather ...

**Energy & Energy Conservation**
Energy Use, Fossil Fuel, Solar, Wind ...

**Environment & Environmental Quality**
Air/Water/Noise Quality, Cleanup, Climate Change ...

**Health & Medicine**
Disease, Health Care, Nutrition, Mental Health ...

**Math, Physics, & Chemistry**
Astrophysics, Chemicals, Mathematical Modeling ...

**Natural Resources & Conservation**
Ecosystems, Energy Resources, Forest Science, Mining

**Science Education**
Homework Help, Teaching Aids, Science Internships ...

## Biotechnology Varieties

The National Agricultural Statistics Service conducts the June Agricultural Survey in all States each year. Randomly selected farmers across the United States were asked if they planted corn, soybeans, or upland cotton seed that, through biotechnology is resistant to herbicides, insects, or both. Conventionally bred herbicide resistant varieties are excluded. Insect resistant varieties include only those containing *bacillus thuringiensis* (Bt). The Bt varieties include those that contain more than one gene that can resist different types of insects. Stacked gene varieties include only those containing biotech traits for both herbicide and insect resistance. The States published individually in the following tables represent 85 percent of all corn planted acres, 88 percent of all soybean planted acres, and 93 percent of all upland cotton planted acres.

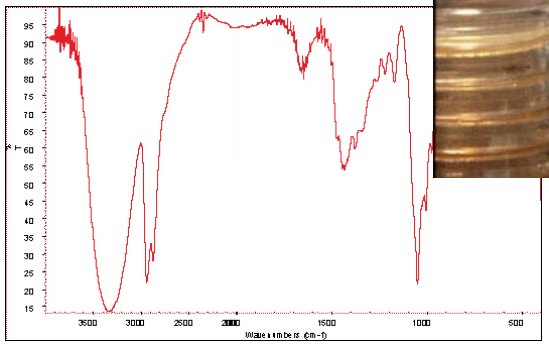
# Search for Insect Resistant Corn Varieties % in South Dakota

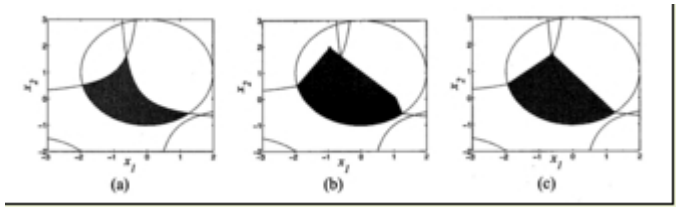
**Corn: Biotechnology Varieties by State and United States, Percent of All Corn Planted, 2007-2008**

| State | Insect Resistant (Bt) | | Herbicide Resistant | |
|---|---|---|---|---|
| | 2007 | 2008 | 2007 | 2008 |
| | *Percent* | *Percent* | *Percent* | *Percent* |
| IL | 1[illegible] | [illegible] | 15 | 15 |
| IN | 1[illegible] | [illegible] | 17 | 16 |
| IA | 22 | 1[illegible] | 19 | 15 |
| KS | 25 | 25 | 36 | 30 |
| MI | 19 | 15 | 22 | 24 |
| MN | [illegible] | [illegible] | 32 | 29 |
| MO | 3[illegible] | [illegible] | 19 | 21 |
| NE | 31 | 27 | 23 | 24 |
| ND | 29 | 24 | 37 | 34 |
| OH | 9 | 12 | 12 | 17 |
| SD | [illegible] | [illegible] | 34 | 30 |
| TX | [illegible] | [illegible] | 37 | 31 |
| WI | 19 | 14 | 23 | 26 |
| Oth Sts [1] | 20 | 20 | 33 | 32 |
| US | 21 | [illegible] | 24 | 23 |
| | Stacked Gene Varieties | | All Biotech Varieties | |

# Before the article

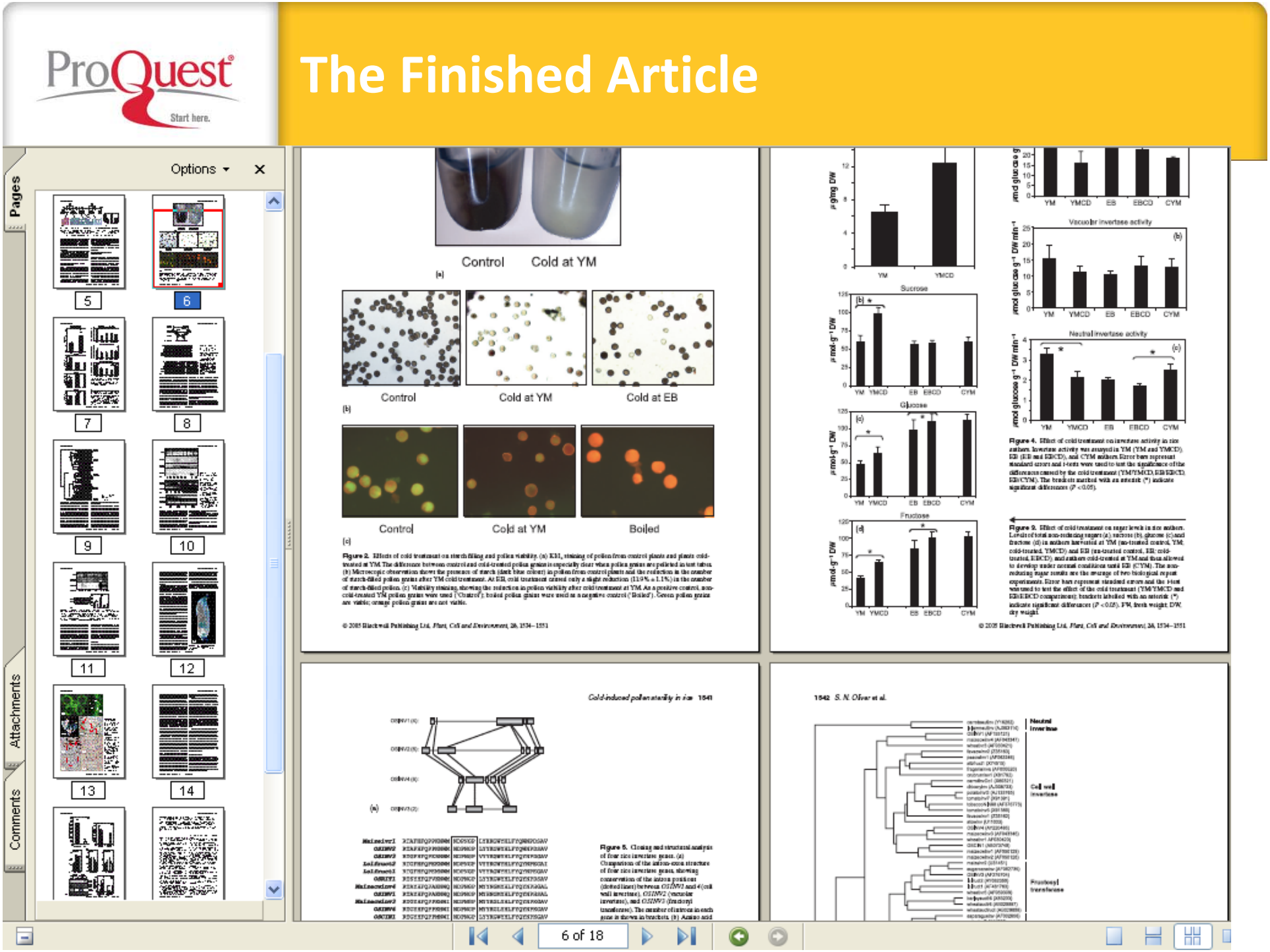
A highly non-linear visual environment



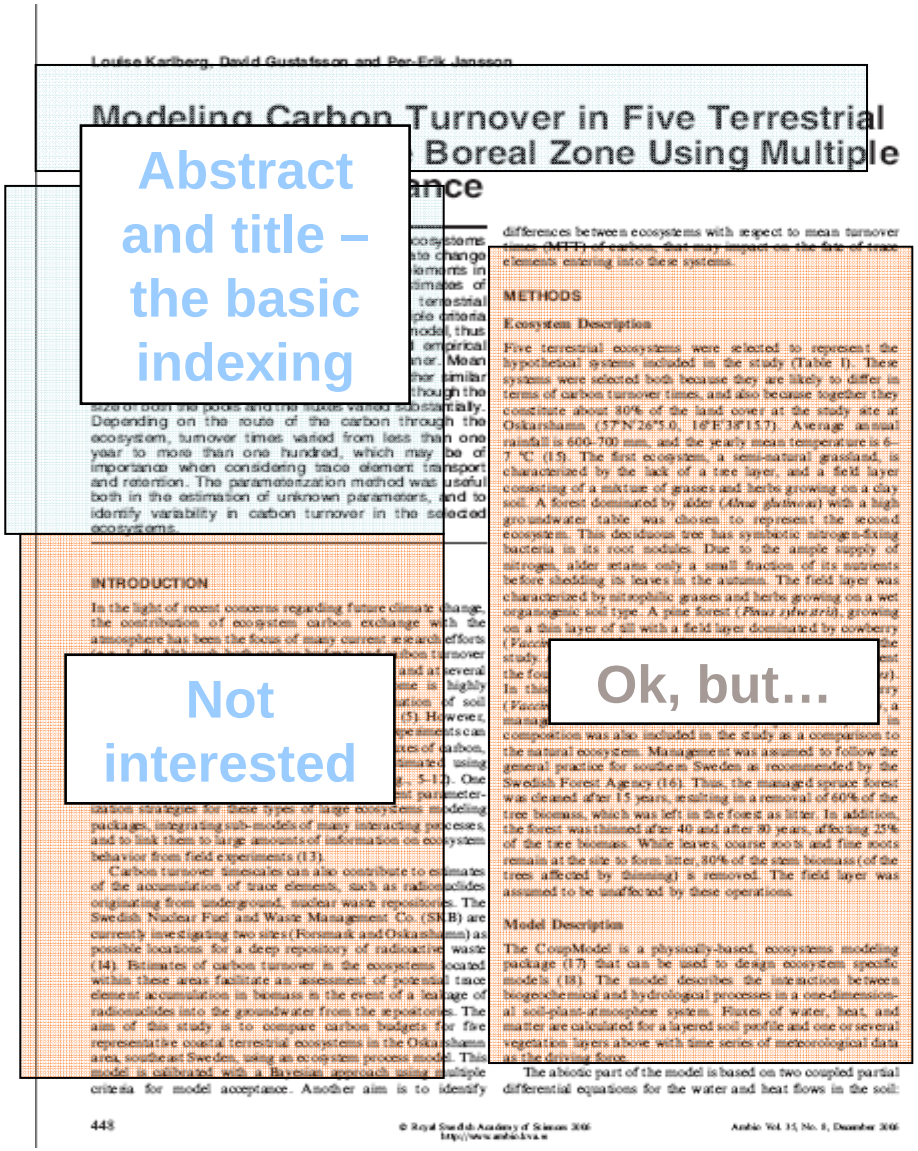



- Months in the field or lab
  - resulting in images, tables, graphs
- Aggregation of information from other articles
  - figures are often used to assess relevance

# The Finished Article



# Article Level Indexing



Louise Karlberg, David Gustafsson and Per-Erik Jansson

Modeling Carbon Turnover in Five Terrestrial … Boreal Zone Using Multiple … ance

Abstract and title – the basic indexing

Not interested

Ok, but…

INTRODUCTION

METHODS

Ecosystem Description

Model Description



© Royal Swedish Academy of Sciences 2006
http://www.ambio.kva.se

We cannot search this

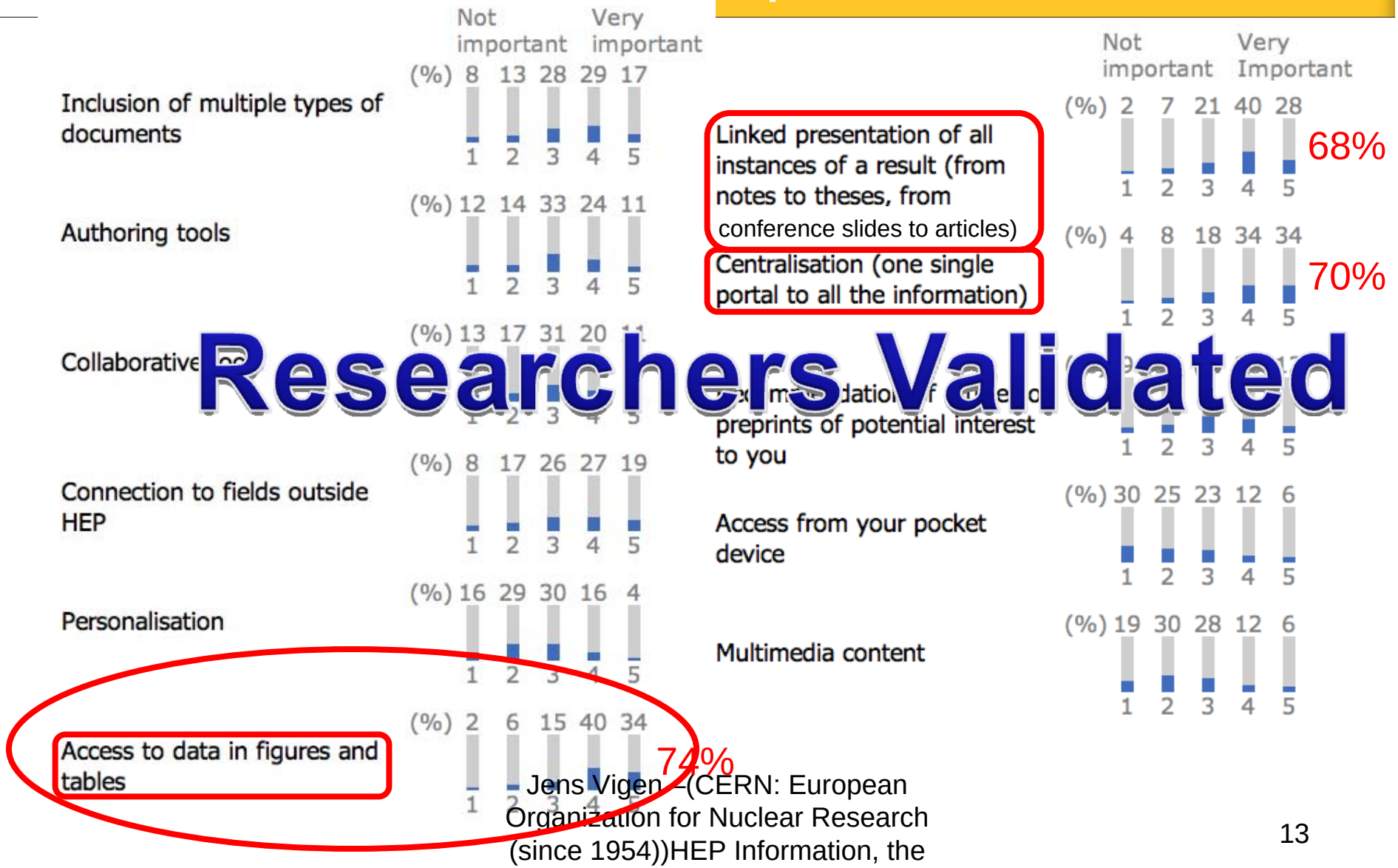
We cannot search this

Parameterization and Model Application



© Royal Swedish Academy of Sciences 2006
http://www.ambio.kva.se

# Imagine your ideal scientific information system in five years…Which additional features will be important?




**Researchers Validated**

Jens Vigen –(CERN: European Organization for Nuclear Research (since 1954))HEP Information, the Bloomsbury Conference 2008

# Findings: End-user Evaluation

- Table & figure searching provides:
  - Higher precision searching
  - Component visibility assists article-level relevance judgment
  - Direct access to high-quality components saves users' time

**Researchers Validated**

THE UNIVERSITY of TENNESSEE School of Information Sciences

UIC UNIVERSITY LIBRARY UNIVERSITY OF ILLINOIS AT CHICAGO

LIBRARYJOURNAL

ProQuest Start here.

**Deep Indexing**:
A New Approach to Searching Scholarly Literature

- **ProQuest Findings:**
  - **Figures and tables represent the distilled essence of research – the closest thing to raw datasets**
  - **Tables and Figures often invisible using traditional article-level and full-text searching**
    - **False Hits when searching for specific tables and figures**
    - **Large Result Sets To Browse**
  - **Establish the relevance of the related data and findings – direct links to document after qualification**

# What is Deep Indexing?

## Deep Indexing Of Data Summaries Makes Objects Visible And Retrievable

- Identification of tables and figures (images, charts, maps, etc.) found within a document
- Extraction of critical data and information within and surrounding each table or figure (including full caption) to provide indexing

# Related Projects

- TableSeer (CiteSeer) - 2006+
  - Automatic extraction of table metadata from documents, Pennsylvania State University
- BioText Search Engine - 2007+
  - PubMed Central documents; functionally similar to Illustrata, but all open access. University of California, Berkeley
- ORE (Object Reuse & Exchange) Aggregation - 2007+
  - Data curation prototype connecting data deposition with the publishing process (links datasets to articles), Johns Hopkins University
- DeLIver - 1996-2000
  - Full text of 50 journals in engineering, physics, and computer science, University of Illinois at Urbana-Champaign

# BioText Search Engine

The system indexes all open access articles available at PubMed Central. New articles are indexed daily. The current collection consists of more than **300 journals**, **40,000 articles**, **100,000 figures**, and **60,000 tables**.

BioText SEARCH ENGINE

Home | About BioText | Contact Us

mg/l

Search

**Search Over:** Full Text & Abstracts  Figure Captions (List)  Figure Captions (Grid)  Tables  **Sort By:** Relevance

Results **1-20** of **320**  << Prev **1** 2 3 4 5 Next >>  **TABLES EXPANDED**

**Review of methodologies and a protocol for the Agrobacterium-mediated transformation of wheat**

Jones, H., Doherty, A., Wu, H. (2005) *Plant Methods*.

TABLE CAPTION

Table 3. Composition of medium **MG/L**

| Component | /litre |
|---|---|
| Mannitol | 5 g |
| **L**-Glutamic acid | 1 g |
| $KH_2PO4$ | 250 **mg** |
| NaCl | 100 **mg** |
| $MgSO_4 \cdot 7H_2O$ | 100 **mg** |
| Tryptone | 5 g |
| Yeast extract | 2.5 g |

Show Full Table (10 rows)

# Google Images

•Needs ”labelling” to retrieve anything sensible:

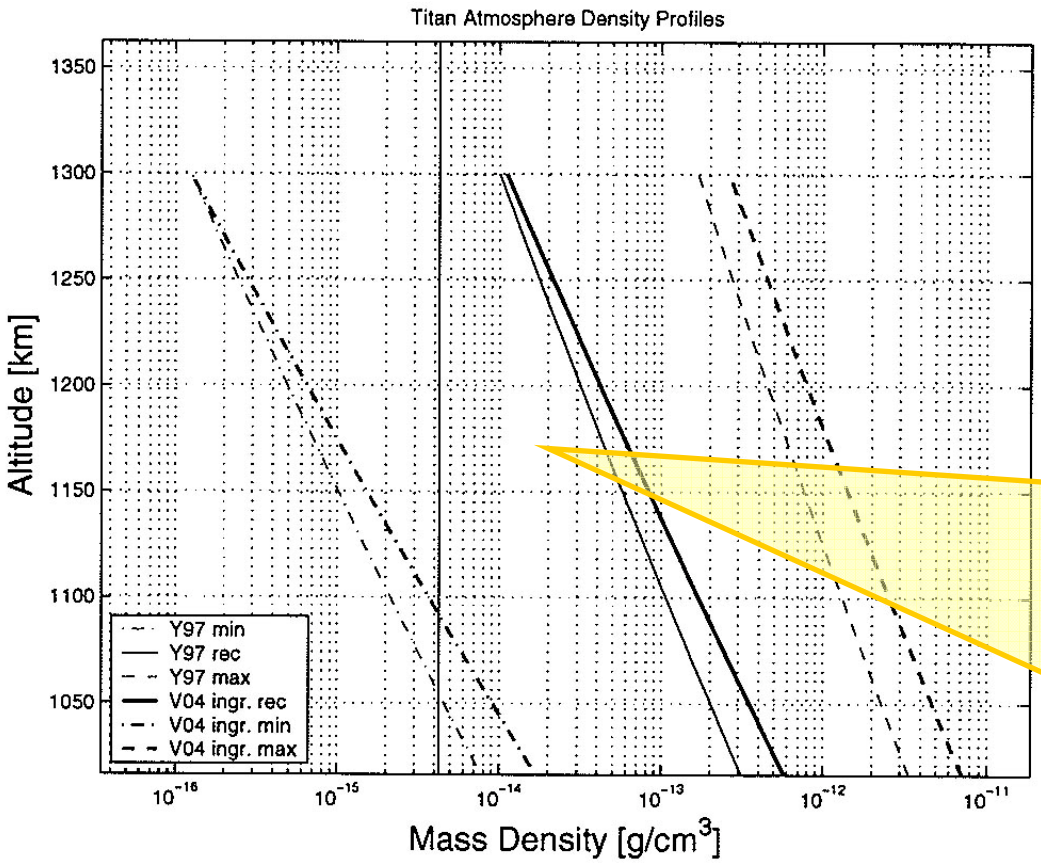
- •Make sure the page topic corresponds to the search term
- •Make an image of standard size{like 800 x 600 }
- •Make the image in .jpg or .gif format.{Not necesarry, but recommended }
- •Use the search term in the alt attribute, the title attribute and make sure that you have included the width and height declarations

•Or the use of their new “labeller”:

*You'll be randomly paired with a partner who's online and using the feature. Over a 90-second period, you and your partner will be shown the same set of images and asked to provide as many labels as possible to describe each image you see. When your label matches your partner's label, you'll earn some points and move on to the next image until time runs out.*

**Does not search within the tables and figures and other images, just in the labelling surrounding them**

# CSA Illustrata- First commercial deep indexing project

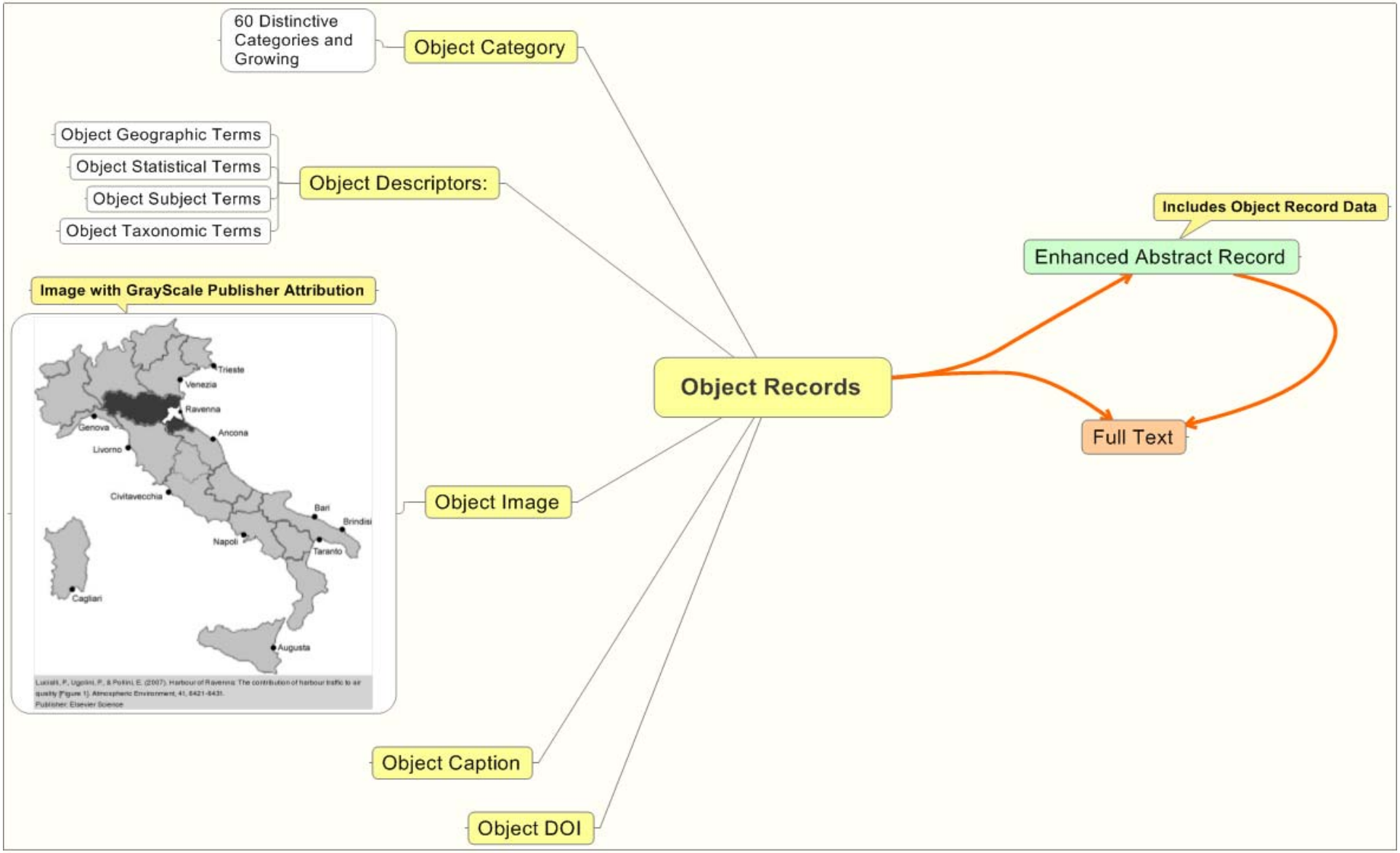



Kazeminejad, B., Lammer, H., Coustenis, A., Witasse, O., Fischer, G., Schwingenschuh, K., et al. (2005). Temperature variations in Titan's upper atmosphere: Impact on Cassini/Huygens [Figure 4]. Annales Geophysicae, 23, 1183-1189.
Publisher: European Geophysical Society

CSA Illustrata makes it possible to search indexed:

- Graphs
- Illustrations
- Maps
- Photographs
- Tables
- Transmission/emission images

# Deep Indexing



# Contextual.....

Please log in to My Research

Logout | Quick Search | Advanced Search | Search Tools | Browse

0 Marked Records | Search

Help & Su

All | Published Works | Tables & Figures | Scholars | Funding | Calls for Papers | Data Curves

Add Row | Remove Row

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | ( fish farm | or | | or | | ) | Anywhere |
| and | ( environment* | or | pollution | or | | ) | Anywhere |
| and | ( | or | | or | | ) | Anywhere |

Search Tips: e.g., wildcar*, exact phrase; use Keywords for a single search of Title, Abstract, Descriptors

Search | Clear

**Now Selected:** CSA Illustrata: Natural Sciences

**Change:** — Subject Area — or Specific Databases

**Date Range:** Earliest to 2008

**More Search Options:** ( Hide )

**CSA Illustrata: Natural Sciences Indexes & Limits**

| | | |
|---|---|---|
| Author | | Browse Author |
| Journal Name | | Browse Journal Name |
| Category | | Browse Category |
| Object Descriptors | | Browse Object Descriptors |

☐ Predictive Model

Category ☐ Graph ☐ Illustration ☑ Map ☐ Photograph ☐ Table ☐ Transmission/Emission Image

# Relevant.....

Quick Search Results - Windows Internet Explorer

http://www-mi9.csa.com/ids70/results.php?id=2&rt=objects&SID=5805f6813987c1c23a3bb246e2b11398&mark_id=search%3A2%3A7%2C0%2C25#resource_tabs

Quick Search Results

10. **Database** CSA Illustrata: Natural Sciences

**Image File** Table 2. Full Image | Full-Text Linking | SFX

| Production sector | Energy required (GJ) | Energy generated (GJ) |
|---|---|---|
| **Sugarcane (Brazil)** | | |
| Agricultural sector | 35.98 | – |
| Industrial sector | 3.63 | 155.57 |
| Distribution | 2.82 | – |
| Total | 42.43 | 155.57 |
| **Corn (United States)** | | |
| Agricultural sector | 22.08 | – |
| Industrial sector | 41.60 | 71.44 |
| Distribution | 1.34 | – |
| Total | 65.02 | 71.44 |

*Note:* Values correspond to ethanol production derived from 1 hectare of sugarcane or corn plantation.

Dias de oliveira, M. E., Vaughan, B. E., & Rykiel, E. J. (2005). Ethanol as Fuel: Energy, Carbon Dioxide Balances, and Ecological Footprint [Table 2]. Bioscience, 55, 593-602.
Publisher: American Institute of Biological Sciences

© Get Permissions Copyright Clearance Center

**Caption** Table 2. Energy balance of ***ethanol production*** from sugarcane in Brazil and from corn in the United States.

**Category** Table

**Title** *Ethanol* as Fuel: Energy, Carbon Dioxide Balances, and Ecological Footprint

**Author** Dias de oliveira, M.E.; Vaughan, B.E.; Rykiel, E.J.

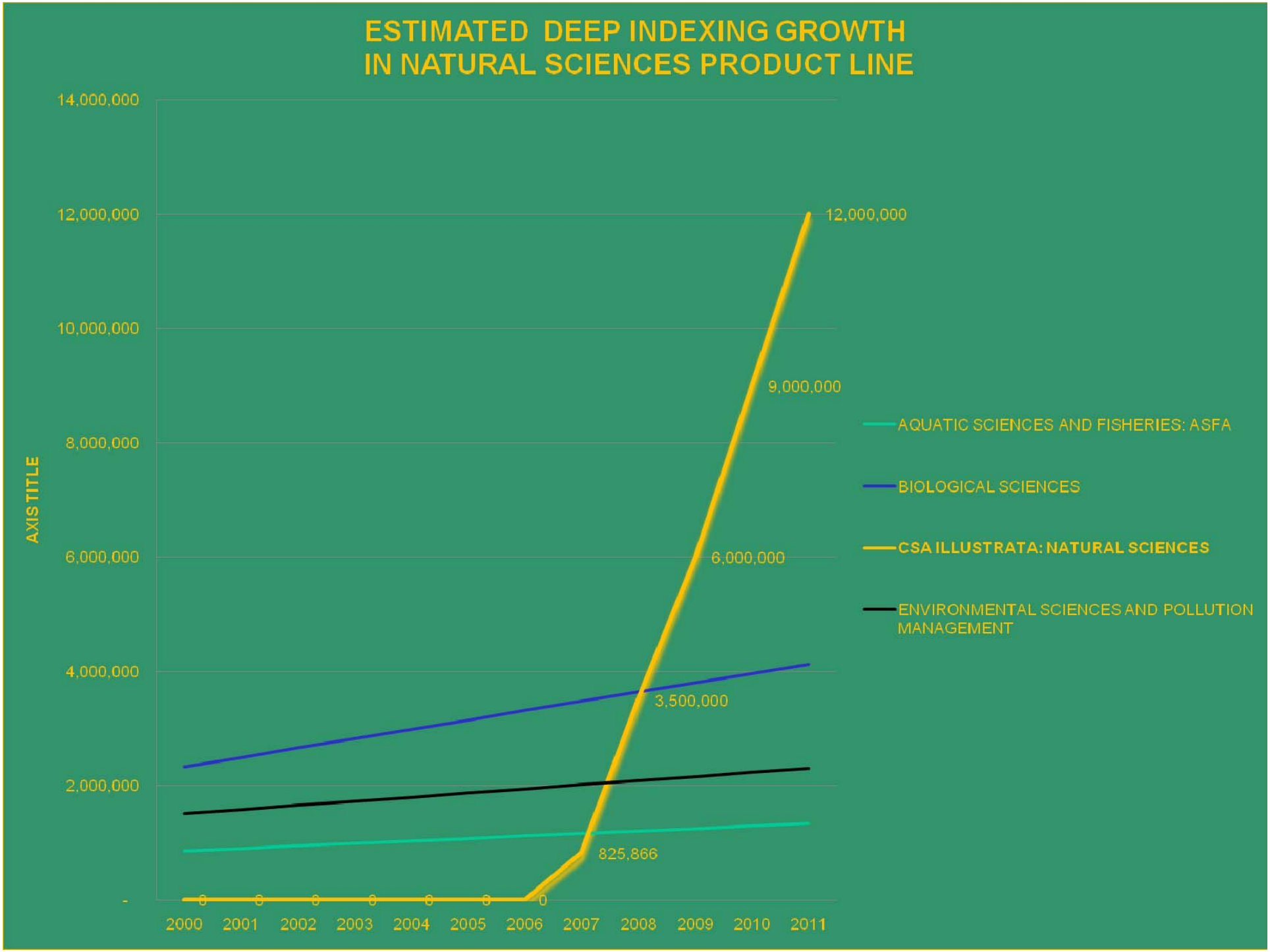
**Source** Bioscience, Vol. 55, No. 7, pp. 593-602. Jul 2005.

**Object Descriptors** **Object Subject Terms** Energy balance of ***ethanol production***; Energy generated GJ; Energy required GJ; ***Production*** sector;

**Object Geographic Terms** Brazil

**Publisher** American Institute of Biological Sciences, 1444 Eye St. N.W., Suite 200, Washington, DC 20005

Local intranet 100%


ESTIMATED DEEP INDEXING GROWTH
IN NATURAL SCIENCES PRODUCT LINE
AXIS TITLE
14,000,000
12,000,000
10,000,000
8,000,000
6,000,000
4,000,000
2,000,000
-
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
0
0
0
0
0
0
0
825,866
3,500,000
6,000,000
9,000,000
12,000,000
AQUATIC SCIENCES AND FISHERIES: ASFA
BIOLOGICAL SCIENCES
CSA ILLUSTRATA: NATURAL SCIENCES
ENVIRONMENTAL SCIENCES AND POLLUTION MANAGEMENT

# New Object Types Tech

## Graph Types

- Control Chart
- Diffraction Pattern
- Gantt Chart
- Phase Diagram

## Diagram Types

- Pole Figure
- Spectrum

## Table Types

- Truth Table

## Illustration Types

•Architectural Design or Blueprint
•Floor Plan
•Mechanical Design or Blueprint
•Floor Plan
•Organizational Chart
•Schematic – Circuit Diagram

# Questions for the Group

- What are the most important sources for objects that may be hidden?
  - Government Users
  - Research Community at Large
- If you start a deep indexing project, what Resources will you need?
    - Document Acquisition

# Creation of the Tables & Figures Index

## Article Acquisition

- Hardcopy
- Scan
- PDF *image*
- OCR
- PDF *text*
- XML *or variant*

## Image Processing

- Manual Image Zoning
- Automated Image Extraction

Machine-Assisted Indexing:

*Subject, Taxonomic; Geographic, Statistical*

Manual Indexing

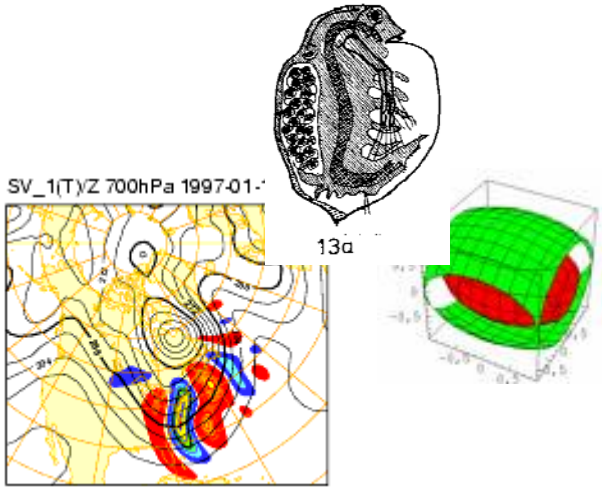
Indexing Review

# Deep Indexing of Objects

## Categorization

- **What "type" of object is represented... a map, a graph?**

…
*Figure — Graph — 3D Surface Plot*
*Figure — Graph — Histogram/Bar Chart*
*Figure — Illustration — Gene/Protein – Maps & Sequences*
*Figure — Illustration — Molecular Structure*
*Figure — Map — Bathymetric Map*
*Figure — Map — Topographic Map*
*Figure — Photograph — Satellite Image*
…




## Addition of Descriptors

- Terms from the dependent and independent variables
- What are the units?
- Subject, taxonomic, geographic terms from caption and article text
- What statistical techniques are used for analysis?

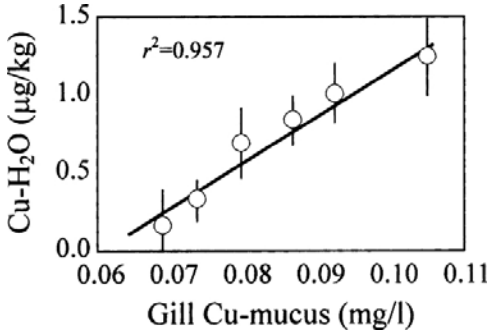
# Deep Indexing of Research Material

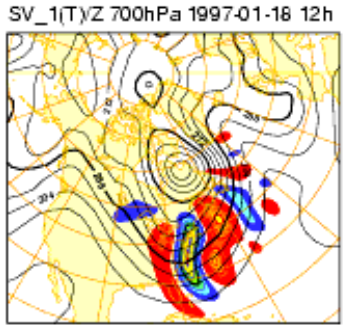
## Selection of other relevant data

May help researcher filter search results

Include:

- Presence of a predictive model
- Indication of data display in color
- Number of figure panels



# To Learn More

- For more information:

  http://info.csa.com/csaillustrata

  - Library Journal Webinar
  - Whitepaper
  - Reviews

- Stay in Touch…Network….

  mark.hyer@proquest.com

  Phone 301-961-6704