



# Text Mining and Publishers

Policy Changes

Presented By Bart Wacek

Date April 7, 2014

## New Policy as of January 2014

*In December, 2013, Elsevier and 12 other publishers, signed an International Association of Scientific, Technical and Medical Publishers declaration expressing a 'commitment to the facilitation of text and data mining for non-commercial scientific research.'*

- **We have updated our TDM policy which enshrines text- and data-mining rights in our standard ScienceDirect subscription agreement for academic customers.**
- **We have ensured all scientific journal content held by Elsevier is now in a mine-able format**
- **We have made all scientific journal content accessible from our platforms via an API which provides full text XML**

## Text and Data Mining Policy

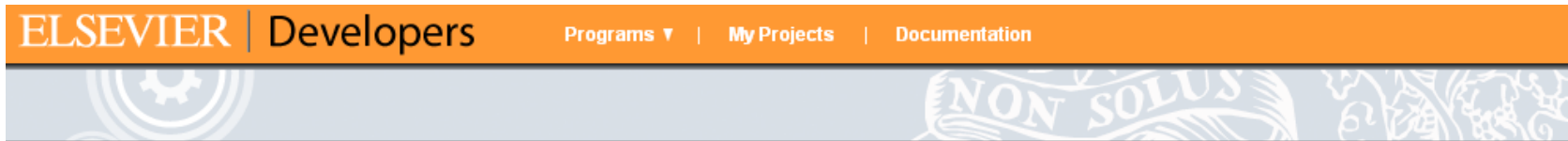
- **Academic subscribers:**
  - Researchers at these institutions can text mine subscribed content on ScienceDirect for non-commercial purposes, via the ScienceDirect API's. Access is granted to faculty, researchers, staff and students at the subscribing institution. No additional charge.
- **Access is via the ScienceDirect APIs only**
  - 1. Output can contain "snippets" of up to 200 characters of the original text
  - 2. Licensed as CC-BY-NC
  - 3. Includes DOI link to original content
- **Non-subscriber access:**
  - Requests are considered on a case-by-case basis, and can be made by emailing: [UniversalAccess@elsevier.com](mailto:UniversalAccess@elsevier.com)

## How to Gain Access

- **For Academic subscribers**
  - Once your institutional agreement has been updated to allow text-mining access, individual researcher access is an automatic process, managed through our developer portal. Researchers will need to follow three steps:
    - Register their details using the online form on the developer's website
    - Agree to our Text Mining conditions via a "click-through" agreement
    - Receive an API token that will allow you to access ScienceDirect content (delivered in an XML format suitable for text mining)

## Developer Page

<http://www.developers.elsevier.com/cms/index>



### The Elsevier API Program

Elsevier offers public APIs for the following products:

- **ScienceDirect** ([www.sciencedirect.com](http://www.sciencedirect.com)) and **Scopus** ([www.scopus.com](http://www.scopus.com)). These APIs are documented on the site you're on now.
- **Mendeley** ([www.mendeley.com](http://www.mendeley.com)). These APIs are documented on <http://dev.mendeley.com>

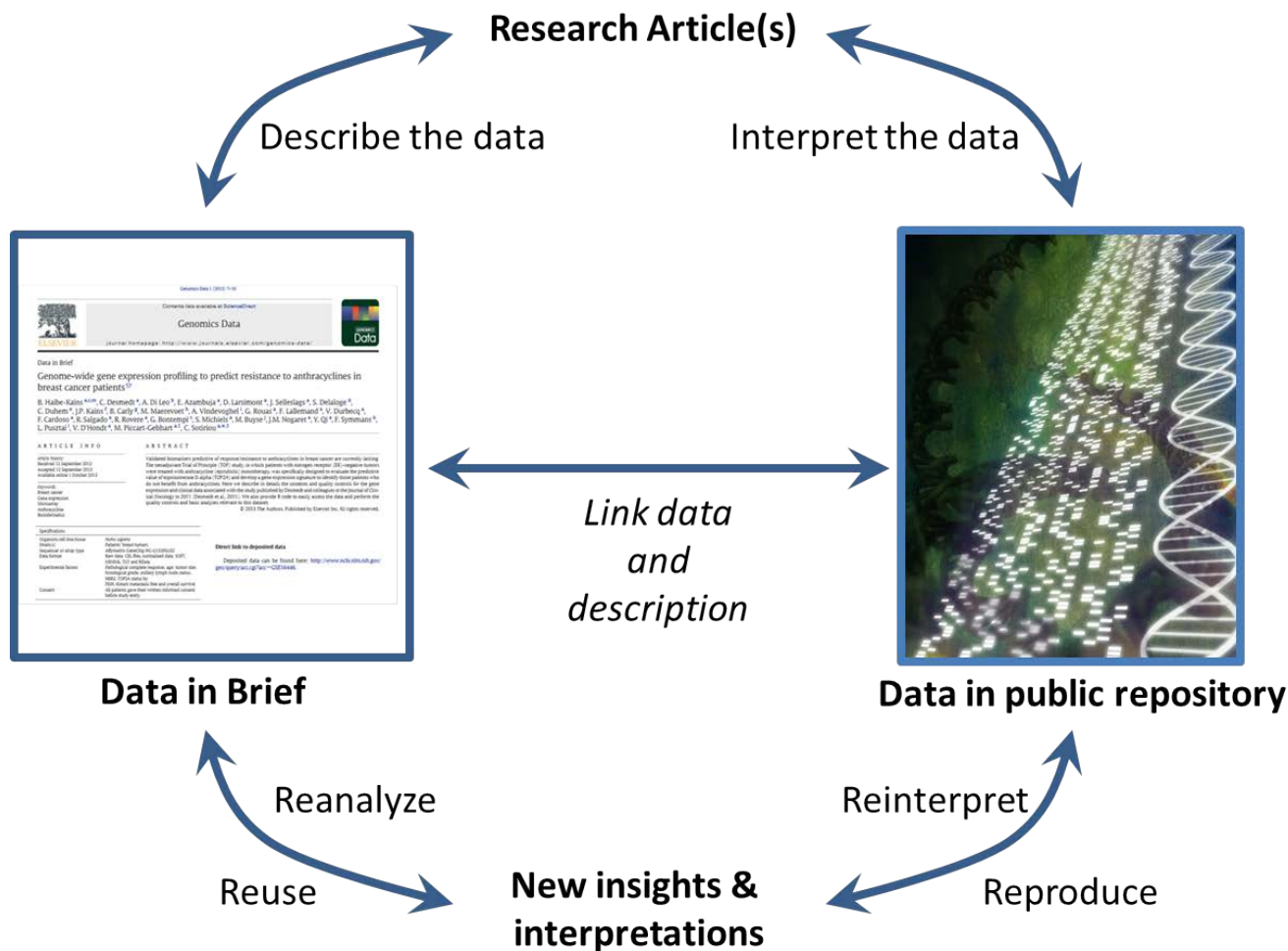
**For getting access to the ScienceDirect/Scopus APIs, take these steps:**

1. Look at some of the common use cases for integrating with the ScienceDirect and Scopus APIs, listed below to see if yours is included. Each use case has pointers to specific documentation and policies.
2. If your use case is not listed, you can find our policy overview [here](#) and general technical API documentation [here](#).
3. Register for API access on the [My Projects page](#):
  - Use the "Register a new site" option if you need API access for something *else* than text mining Elsevier's full-text content.
  - Use the "Register a new text mining project" option if you need API access for text mining Elsevier's full-text content. If you do not see this option, [contact us](#).

## Miscellaneous Points on Policy

- **Figures and images included in metadata**
  - Reference books, too, on ScienceDirect
- **Non-commercial research purposes**
- **No third-party users on behalf of researchers**
- **No crawlers for PDF**
  - Still human only
  
- **More collaboration across the entire publishing industry**
  - CrossRef Prospect
  - Common API and common license framework
- **Ultimate goal of 'enhanced' text mining**
  - Tool box in the cloud

# Micro Article



## Genomics Data (1<sup>st</sup> of many disciplines)

- **Data in Brief articles:**
  - Thoroughly describe data, facilitating reproducibility.
  - Make deposited genomic data easier to find.
  - Increase traffic towards associated research articles and data, leading to more citations.
  - Open up doors for new collaborations.
  
- **“Toolbox in the Sky”**
  - In Development
  - Data in Brief articles describe genomic data thoroughly so that the data can be reused and reanalyzed. Whether you’ve already published a research article with your data or have just uploaded your data into a public repository, make the most of that data and write a quick, Open Access Data in Brief article.