# 1.5B Data Points trapped in International Repository - Daring Rescue in Progress!

**Simon N. Twigger[1,2], Joey F. Geiger[2], Jennifer R. Smith[1], Rajni Nigam[1], Clement Jonquet[3], Mark Musen[3]**
[1] Human and Molecular Genetics Center and [2]: Biotechnology and Bioengineering Center, Medical College of Wisconsin, Milwaukee, WI, USA, [3] National Center for Biomedical Ontology, Stanford University, Palo Alto, CA, USA

**Project data is available at http://gminer.mcw.edu**

RGD Rat Genome Database
http://rgd.mcw.edu

THE NATIONAL CENTER FOR BIOMEDICAL ONTOLOGY

MEDICAL COLLEGE OF WISCONSIN

Has anyone done any expression studies using congenic rats?

What tissue is this gene expressed in?

What expression data is known for SD (aka SD/NHsd, Harlan Sprague Dawley, Sprague Dawley) rats?

Are any of these genes associated with my phenotype?

Has this gene been seen in the brain?

What rat expression studies have been done on Mammary Cancer(aka breast neoplasms/breast cancer/cancer of the breast, breast carcinoma...)?

```
INDEX
Brain            11 hits
  Forebrain       5 hits
  Hindbrain       3 hits
  Amygdyla        3 hits

SS rats          23 hits
  SS/Jr           9 hits
  SS/NHsd        10 hits
  S.R(9)x3a       4 hits
  ...
```

Expression data is at the core of many critical questions asked by researchers.

NCBI's Gene Expression Omnibus has a lot of data but results are very detailed, centered on experiments, not genes.

Lots of valuable information is encoded in the text of the records but is not easy to access or repurpose.

Can we get more information out of GEO by identifying information in the text and creating an Index of the GEO records? Can we link this index to the billions of data points stored in GEO to extract more information?

THE NATIONAL CENTER FOR BIOMEDICAL ONTOLOGY

Mouse Anatomy Ontology

Rat Strain Ontology

http://bioontology.org

Using ontologies and the ontology structure allows queries not possible using other search approaches

We are exploring linking our annotations and GEO expression data to derive novel gene-to-tissue relationships

Probeset **1369048_at** (Gabrd) was **P**resent in sample GSM132484 which was from **hippocampus**. Can we compile all such results for Gabrd...?

Use Linked Tag Clouds to navigate the GEO data to find GEO records associated with strains or anatomical regions. Tag Clouds also provide a summary of the search results as well as an intuitive navigation aid.

We use NCBO tools and an automated pipeline to match ontology terms to GEO text (concept mapping) and manually review to create the index.

National Center for Biomedical Ontology hosts a wide variety of Ontologies that provide a structure for knowledge in many critical areas