

Title of Project: Development and use of digital methods at the Sociomics Core Facility: Visualisation of Patent information.

Name of Supervisor(s) Paul Oldham and Stephen Hall

Outline of Project.

Building on the success of the 2009 SCF internship this programme provides a young computer programmer with six weeks work experience with social scientists in developing and using digital methods at the Sociomics Core Facility at Cesagen, Lancaster University. The programme focuses on delivery of useful outcomes (i.e. software and tools) within that period as part of training in real world applications. As the internship focuses on programming the internship is task focused.

The internship will focus on two tasks that will enhance methodological development and access to patent data for social scientists and policy-makers.

1. Visualization of World Patent Statistical Database data (PATSTAT) in Google Maps (Research Desktop) and Google Earth

This task will focus on exploiting the availability of a patent database with 60 million records at the Sociomics Core Facility. The task will exploit the availability of address field data for patent applicants in the address field of the World Patent Statistical Database (PATSTAT) at the SCF.

The task will involve.

1. Extracting country code data for a defined dataset from the priority, publication country, patent assignee and inventor for display in Google Maps and Google Earth. This may include - showing networks of the country of origin (priority) such as the United States and the country of publication i.e. India/elsewhere;
2. Where more than one country code is listed in the field code data (i.e. assignee or inventor) it will be useful to visualize as a network;
3. Extracting address field data (i.e. zip code, post code data) from PATSTAT for display in Google Maps/Google Earth. Analysis of the European Patent Office REGPAT dataset (available at the SCF) that uses European Standardized address codes (NUTS codes) will be useful here.

2. Visualizing Technology Trends using the International Patent Classification:

The World Patent Statistical Database contains the world's patent information dating to the 19th Century. That information is classified using the 70,000 classification codes in the International Patent Classification. Classification codes are hierarchical (i.e. C12 for biochemistry and biotechnology then C12N for genetic engineering) and documents are awarded more than one classification code. That means they can be placed in a matrix of co-occurrences. We would like to do two things.

1. To link patent data for specific classification codes (i.e. A61K36 for traditional medicines) to the information in task 1 above so that users can visualize trends for different countries in different areas (i.e. A61K36 against C12N for biotechnology) in google maps/google earth
2. To generate a separate technology landscape map of co-occurring IPC codes that can be accessed at different levels based on a co-occurrence matrix to be created in PATSTAT.

That co-occurrence matrix should include the following components:

- a. Ability to interrogate the hierarchy at different levels (i.e. C12N as sub class or C12NI5/82 as the group for agricultural biotechnology)
- b. Ability to visualize transformations in the landscape over time (i.e. decades) to slice through.
- c. Possible linkages to other code data (i.e. countries or assignees) to identify the main players in an area of the technology landscape.

Outcomes. Please state the likely outcomes of this project.

Online access to PATSTAT data as indicators for use by researchers and policy makers. Online access to patent data visualizations using Google Maps in the Research Desktop and possible separate access in Google Earth.

A publication on the global technology landscape and its evolution over time.