Project Goals

(1) Conduct a **detailed analysis of the information needs** of a representative set of science policy makers including existing data, approaches, and tools.

(2) Develop a **theoretic conceptualization of tasks** relevant to science policy-making that map the needs of policy makers to theoretically grounded and practically valuable processing pipelines that transform data into actionable information.

(3) **Design a prototypical tool, a macroscope**, to see structure, patterns, trends, and outliers in science and technology (S&T) data sets that are too large and complex to be comprehensible to us – just like microscopes and telescopes help us to see things that are too small or too far away. The Macroscope tool development will benefit from the NSF funded **Scholarly Database (SDB)** that provides access to more than 20 million scholarly records, and the **Cyberinfrastructure Shell (CIShell)** which supports the easy plug-and-play of datasets and algorithms and the design of stand-alone tools. Introduce the validated macroscope to a broader audience by means of the **Places & Spaces: Mapping Science** exhibit.
1. Detailed Needs Analysis

A total of 34 science policy makers and researchers at university campus level (8), program officer level (12), and division director level at national, state, and private foundations (10) as well as science policy makers from Europe and Asia (4) were interviewed between Feb. 8th, 2008 and Oct. 2nd, 2008.

Each interview comprised a 40 min, audio-taped, informal discussion on specific information needs, datasets and tools currently used, and information on what a 'dream tool' might look and feel like. There is also a pre-interview questionnaire to acquire demographics and a post-interview questionnaire to get input on priorities.

Data compilation is in progress, should be completed in July 2009, and will be submitted as a journal paper. Some data excerpts are given here.

In the Post-Questionnaire Subjects were asked:

“What are initial thoughts regarding the utility of science of science studies for improving decision making? How would access to datasets and tool speed up and increase the quality of your work?”

Excerpts of answers:

- Two areas have great potential: Understanding S&T as a dynamic system, means to display, visualize and manipulate large interrelated amounts of data in maps that allow better intuitive understanding.
- Look for new areas of research to encourage growth/broader impacts of research--how to assess/ transformative science--what scientific results transformed the field or created a new field/ finding panelists/reviews/ how much to invested until a plateau in knowledge generation is reached/how to define programs in the division.
- Scientometrics as cartography of the evolution of scientific practice that no single actor (even Nobel Laureates) can have. Databases provide a macro-view of the whole of scientific field and its structure. This is needed to make rational decision at the level of countries/states/provinces/regions.
- Understanding where funded scientists are positioned in the global map of science.
- Self-knowledge about effects of funding/ self-knowledge about how to improve funding schemes.
- Ability to see connections between people and ideas, integrate research findings, metadata, clustering career measurement, workforce models, impact (economic/social) on society-interactions between levels of science; lab, institution, agency, Fed Budget, public interests.
- It would be valuable to have tools that would allow one automatically to generate co-citation, co-authorship maps…I am particularly interested in network dynamics.
It would enable more quantitative decision making in place of an "impression-based" system, and provide a way to track trends, which is not done now.

When NSF started SciSIP, I was skeptical, but I am more disposed to the idea behind it now although I still don't have a clear idea what scientific metrics will be.....how they will apply across disciplines and whether it's really possible to predict with any accuracy the consequences of any particular decision of a grant award.

SoS potentially useful to policymakers by providing qualitative and quantitative data on the impacts of science toward government policy goals...ideally these studies would enable policy makers to make better decisions for linking science to progress toward policy goals.

Tracking faculty's work over time to determine what factors get in the way of productivity and which enhance, e.g. course-releases to allow more time--does this really work or do people who want to achieve do so in spite of barriers.

I'm not sure that this has relevance to my decision-making. There is a huge need for more reliable data about my organization and similar ones, but that seems distinct from data and tools to study science.

It would assist me enormously.

Help to give precedents that would rationalize decisions--help to assess research outside one's major area. Ways of assessing innovation, ways of assessing interactions (among researchers, across areas, outside academia).

It would allow me to answer questions from members of congress provide visual presentations of data for them.

Very positive step--could fill important need in understanding innovation systems and organizations.

2. Conceptualizations of Science


3. Macroscope Tool
Benefits from and extends the Scholarly Database at IU

“From Data Silos to Wind Chimes”

- Interlink creators, data, software/tools, publications, patents, funding, etc.
- Create public databases that any scholar can use. Share the burden of data cleaning and federation.

<table>
<thead>
<tr>
<th>Dataset</th>
<th># Records</th>
<th>Years Covered</th>
<th>Updated</th>
<th>Restricted Access</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medline</td>
<td>17,764,826</td>
<td>1898-2008</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>PhysRev</td>
<td>398,005</td>
<td>1893-2006</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>PNAS</td>
<td>16,167</td>
<td>1997-2002</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>USPTO</td>
<td>3,710,952</td>
<td>1976-2008</td>
<td>Yes*</td>
<td></td>
</tr>
<tr>
<td>NSF</td>
<td>174,835</td>
<td>1985-2002</td>
<td>Yes*</td>
<td></td>
</tr>
<tr>
<td>NIH</td>
<td>1,043,804</td>
<td>1961-2002</td>
<td>Yes*</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>23,167,642</td>
<td>1893-2006</td>
<td>4</td>
<td>3</td>
</tr>
</tbody>
</table>

Aim for comprehensive time, geospatial, and topic coverage.
Grant-Article Linking

- NIH grant data from CRISP and RaDiUS were linked to Medline papers using the grant information strings in Medline (dirty data using dozens of formats)
- 94% of grant strings were matched with a grant number
- Enables future input-output studies

**Subsequent Analysis From Matches**

- Short grants (1-2 years) produce more papers per year than long grants (3-15 years).
- Data not normalized for grant size.
- Acknowledgement of NIH funding in Medline-indexed articles does seem to be reasonably complete.
- “None” category size consistent with other analyses – these are not “missing NIH” data.
Scholarly Database: Web Interface

Anybody can register for free at https://sdb.slis.indiana.edu to search the about 23 million records and download results as data dumps.
Currently the system has over 100 registered users from academia, industry, and government from over 60 institutions and four continents.

3. Macroscope Tool
Builds on and extends the Network Workbench tool, ultimately be ‘packaged’ as a SciPolicy branded tool.

The Network Workbench (NWB) tool supports researchers, educators, and practitioners interested in the study of biomedical, social and behavioral science, physics, and other networks.
In Feb. 2009, the tool provides more 100 plugins that support the preprocessing, analysis, modeling, and visualization of networks.
More than 40 of these plugins can be applied or were specifically designed for S&T studies.
It has been downloaded more than 18,000 times since Dec. 2006.

http://nwb.slis.indiana.edu/
SciPolicy Studies - Using Open Data and Open Code

See https://nwb.slis.indiana.edu/community  July 1st, 2008
Mapping Science Exhibit – 10 Iterations in 10 years
http://scimaps.org/

The Power of Maps (2005)

Science Maps for Economic Decision Makers (2008)


Science Maps for Science Policy Makers (2009)
Science Maps for Scholars (2010)
Science Maps as Visual Interfaces to Digital Libraries (2011)
Science Maps for Kids (2012)
Science Forecasts (2013)

The Power of Forecasts (2007)

How to Lie with Science Maps (2014)

Exhibit has been shown in 49 venues on four continents. Also at
- NSF, 10th Floor, 4201 Wilson Boulevard, Arlington, VA.
- University of Alberta, Edmonton, Canada, Nov 10-Jan 31, 2009
- Center of Advanced European Studies and Research, Bonn, Germany,