

**Draft: FOR DISCUSSION**  
**Reports to Science Agencies**

**Criteria:**

- Must create automated reporting that can be updated on a flow basis
- Must be tailored to specific science agency needs
- Must be subsettable to a specific timeframe, program, discipline, or type of investment

**General:**

1. Total investments (source: administrative records)
  - a. Total \$
  - b. By disciplines
  - c. Historical patterns
  - d. Year (length of investment)
  - e. By geographical region
  - f. Investment Type (individual PI, group PI, center, 1 year, new PI, etc)
2. Economic
  - a. Job Creation and retention (source: administrative records)
    - i. Total
    - ii. By Occupation
    - iii. Historical pattern
3. STEM education undergraduate and graduate students (source: administrative records)
  - a. By discipline
  - b. Position retention/creation
  - c. Career trajectories\*
4. Innovation
  - a. Patents (source: PTO; confirmed by PIs)
  - b. Patent Applications (source: PTO; confirmed by PIs)
  - c. Co-location clustering (source: patent application addresses, university addresses, PI addresses, Fed Lab addresses)
  - d. Start-up/IPOs\*
  - e. Economic rents\*
  - f. Licensing returns\*
5. Advancing Knowledge
  - a. Publications of PIs (source: webscraping; confirmed by PIs)
    - i. Publication Outlets
    - ii. Publication Number/counts
    - iii. Share of publications relative to reference (university, discipline, agency, etc)
    - iv. Discipline published in/Field analysis (identifies new fields, fostering novel research areas)
  - b. Citations of PIs (source: webscraping; confirmed by PIs)

- i. Relative citation impact (discipline specific Journal impact factors) –
  - ii. Highly cited publication
- 6. Use/Absorptive Capacity (hardest to measure, includes both public absorptive capacity and “decision maker” absorptive capacity \*)
  - a. Co-author networks (richness metric, visualization co-author map) \*
  - b. Download rate hit rates\*
  - c. Guideline use\*
  - d. Standard/model/method adoption by industry (i.e. utilities)\*
  - e. Press/media coverage\*
  - f. Inclusion in curricula (source: university course offering, Texas textbooks)\*
  - g. Cited in Federal/State regulations (source: federal dockets, federal registries) \*
  - h. Inclusion as judicial justification (source: court dockets) \*
  - i. Included in Congressional/State Bills/Laws (source: webscraping) \*
- 7. Other outcomes (engagement of community critical to develop these)
  - a. Health
  - b. Equity
  - c. Safety
  - d. Security
  - e. Infrastructure
  - f. environment

### **Additional Possible Features**

Visualization of job creation (cut by geographic location and discipline)

Visualization of publications (discipline)

Visualization of citations relationships

Visualization of patents and patent applications (cut by geographic location and discipline)

Map of citations/patents/patent applications

Visualization of emerging scientific networks

Visualization of agency investment (disciplines and locations) NOT agency specific shows overlapping areas of activity

### **Possible approach**

Step 1: Using preexisting data and frame

1. Create frame of individuals receiving science funding. This frame should
  - a. Initially use only publicly available data on PIs and coPIs
  - b. Go back as far as possible (ten year window target: 2000 – 2010)
  - c. Subsequently be updated on a monthly basis from administrative data from universities and funding agencies
2. Provide initial match to outcomes using existing data on patents, patent applications and citations, which allows for location and discipline specific outcomes. This would build on existing work prototyped by both science agencies and academic researchers.

3. Create prototype reports for principal investigators, universities and science agencies (see below). All should use the following approaches
  - a. Use the existing frame
  - b. Be structured to be updated on a monthly basis
  - c. Provide visually interesting summaries and updates
  - d. Provide useful and relevant information as determined by users through feedback on reports (format, content, and timing) and reported and future metrics.
4. Post prototypes on data.gov (initially on restricted access site)

**Step 2: Expanding the data sources**

1. Update match to outcomes on flow basis using webscraping technologies and administrative data

**Step 3: Expanding the user base and engaging the community**

1. Create openly accessible data enclave for researcher community
2. Announce the existence of the enclave and create prize for best report in each of the four categories (innocentive type competition)
3. Update data.gov with best reports as rated by the user community