CHANGING WHAT'S POSSIBLE

ARPA-E IMPACTS Overview

Digital Innovation for Energy Innovation

Agenda

- Introduction to ARPA-E and IMPACTS
- Motivation and Approach
- Sample Results
  - Case study: Ascertain impact
  - Portfolio analysis: Operate for impact
  - Natural language processing: Plan for impact
- Discussion

What is ARPA-E?

The DOE's Advanced Research Projects Agency-Energy (ARPA-E):

- Provides **Research and Development** funding for high-risk, high-reward, transformational ideas
- Focuses on technologies that could **fundamentally change** the way we get, use and store energy
- Accelerates energy innovations that will create a more secure, affordable, and sustainable **American energy future**

arpa·e CHANGING WHAT'S POSSIBLE

U.S. DEPARTMENT OF ENERGY

ARPA-E Impact Indicators to date

Since 2009 ARPA-E has provided

$2.4 billion

in R&D funding to more than **875 projects**

166 Projects have attracted more than

$3.3 billion

in private-sector follow-on funding

86 companies formed by **ARPA-E projects**

229 projects have **partnered with other government agencies** for further development

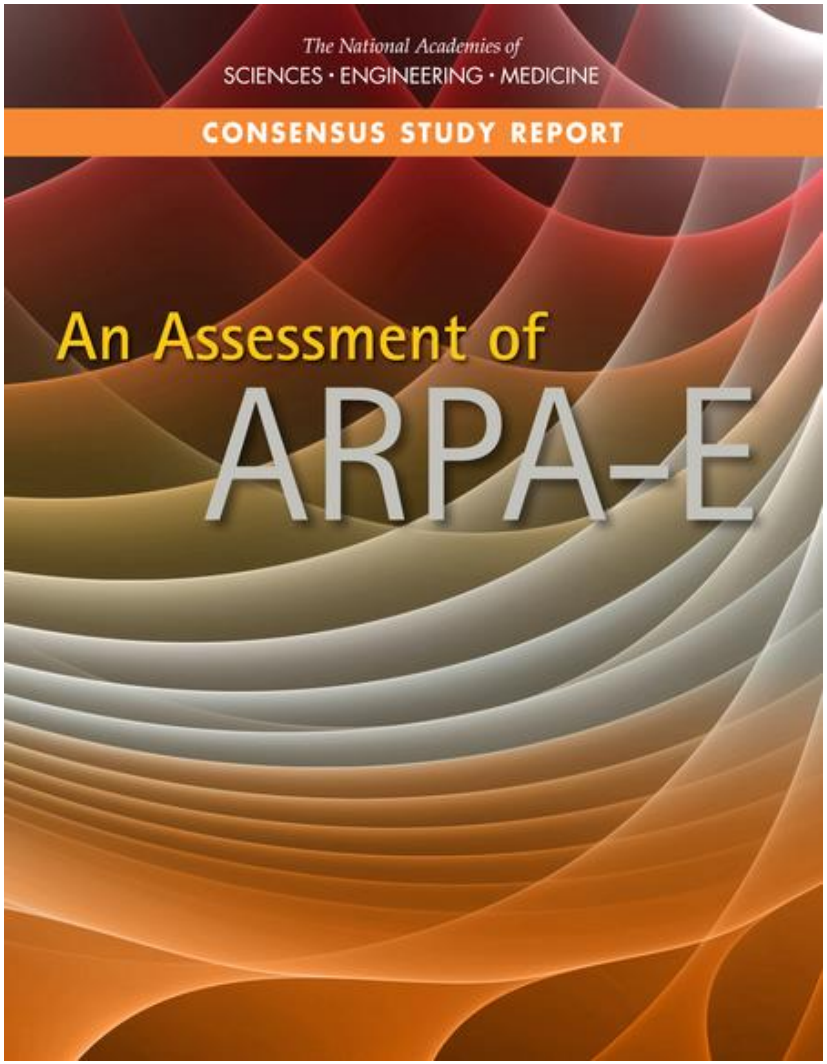
4,021 peer-reviewed **journal articles** from ARPA-E projects

609 patents issued by U.S. Patent and Trademark Office

As of July 2020

… while these indicators provide tangible metrics of mission success, ARPA-E seeks to pioneer the next set of enhanced indicators using advanced analytics

2016 NAS Assessment challenged ARPA-E in this area



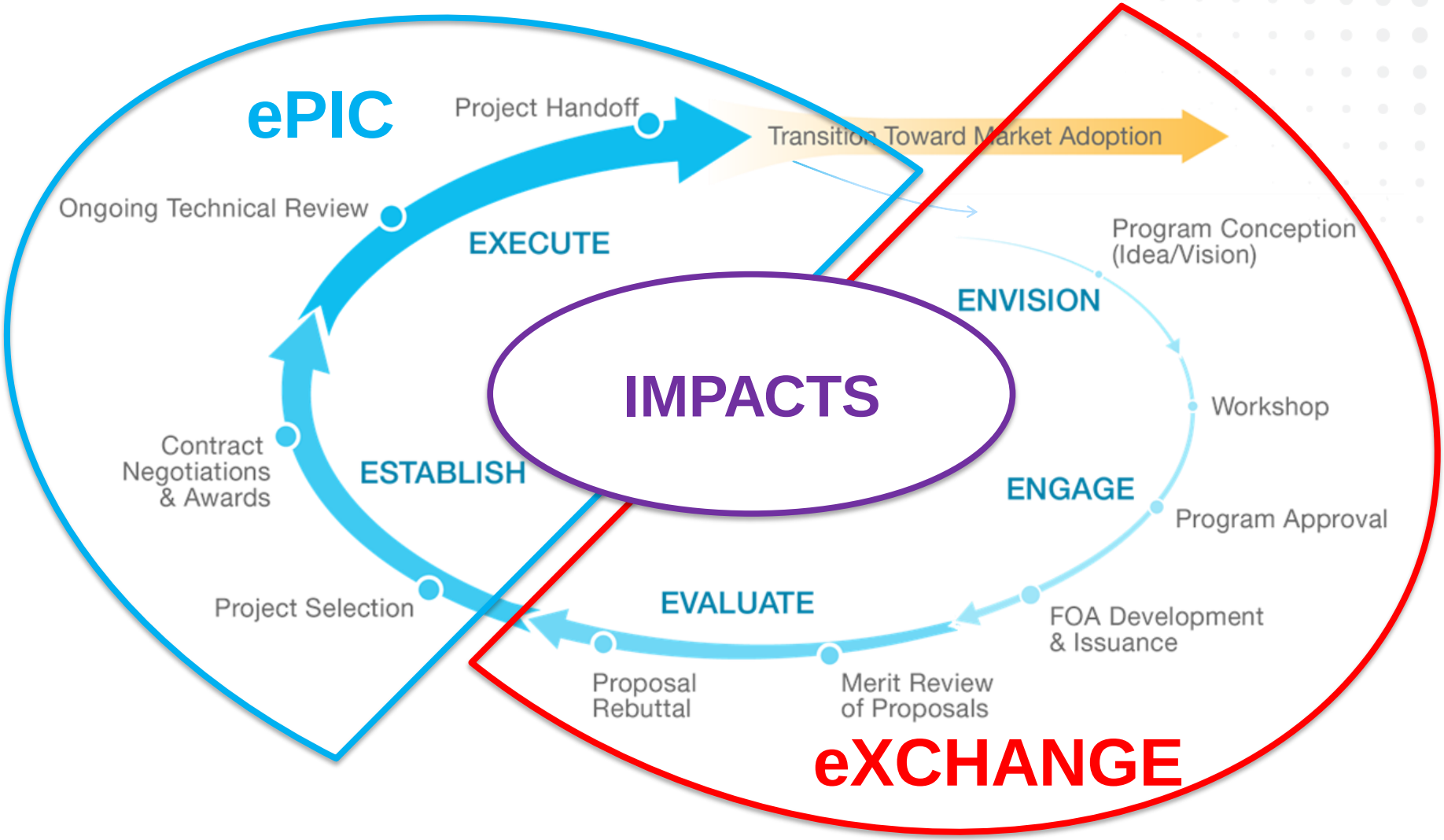
- **Recommendation 4-8:**

  "The ARPA-E director and program directors should develop and implement a framework for measuring and assessing the agency's impact in achieving its mission and goals."

Why is an enhanced framework for evaluating impact necessary?

- Quick access to historical and on-going research
  - Program white space and novelty validation
- “Mining” data to assist with operations
  - Screening for potential investors, reviewers, workshop participants, and recruitment
- Post-award
  - Follow on funding
  - Subject invention, publication and patents
  - “Innovation” level of ARPA-E awards vs. overall sector

ARPA-E R&D Management Lifecycle business processes/systems establish the framework



IMPACTS Capabilities Available Today

- Automatic data Ingestion
  - Python
- NoSQL storage
  - Graph database
  - Document database
  - Structured & un-structured
  - Scalable for future needs

mongo DB · neo4j · python

DATA ANALYSIS

- Search Platform
  - Projects
  - Proposals
- Text Analytics
- Graph Analytics
- Statistical Analysis
- Machine Learning

python · R · Solr · Lingo4G

DATA VISUALIZATION

- Discover hidden relationships in highly connected data
- Identify patterns and clusters
- Turbo-charge communications

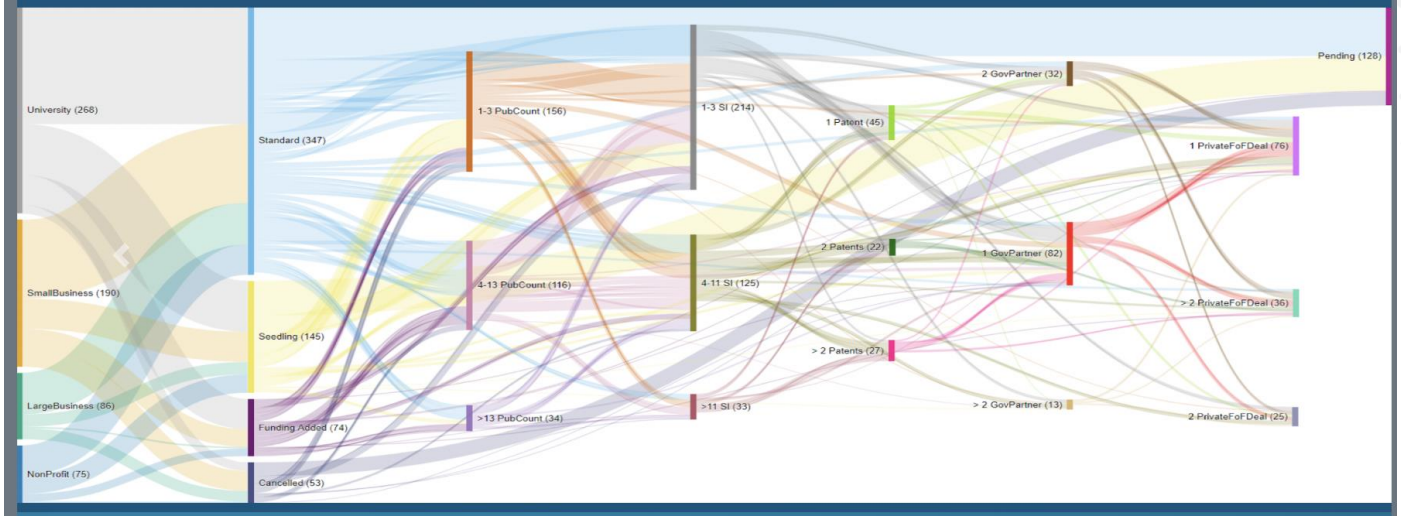
neo4j · R

arpa·e

Home | Getting Started | About | Login

IMPACTS

Indicators Measuring Progress & Change Tracking System



NETWORK ANALYSIS

Graph Platform to map, analyze, store and traverse networks of connected data to reveal invisible contexts and hidden relationships.

TEXT CLUSTERING

Explore topic clusters and identify whitespace in concept papers, full applications, progress reports, and publications

SEARCH PLATFORM

Dynamically search the text of concept papers, full applications, progress reports, and publications

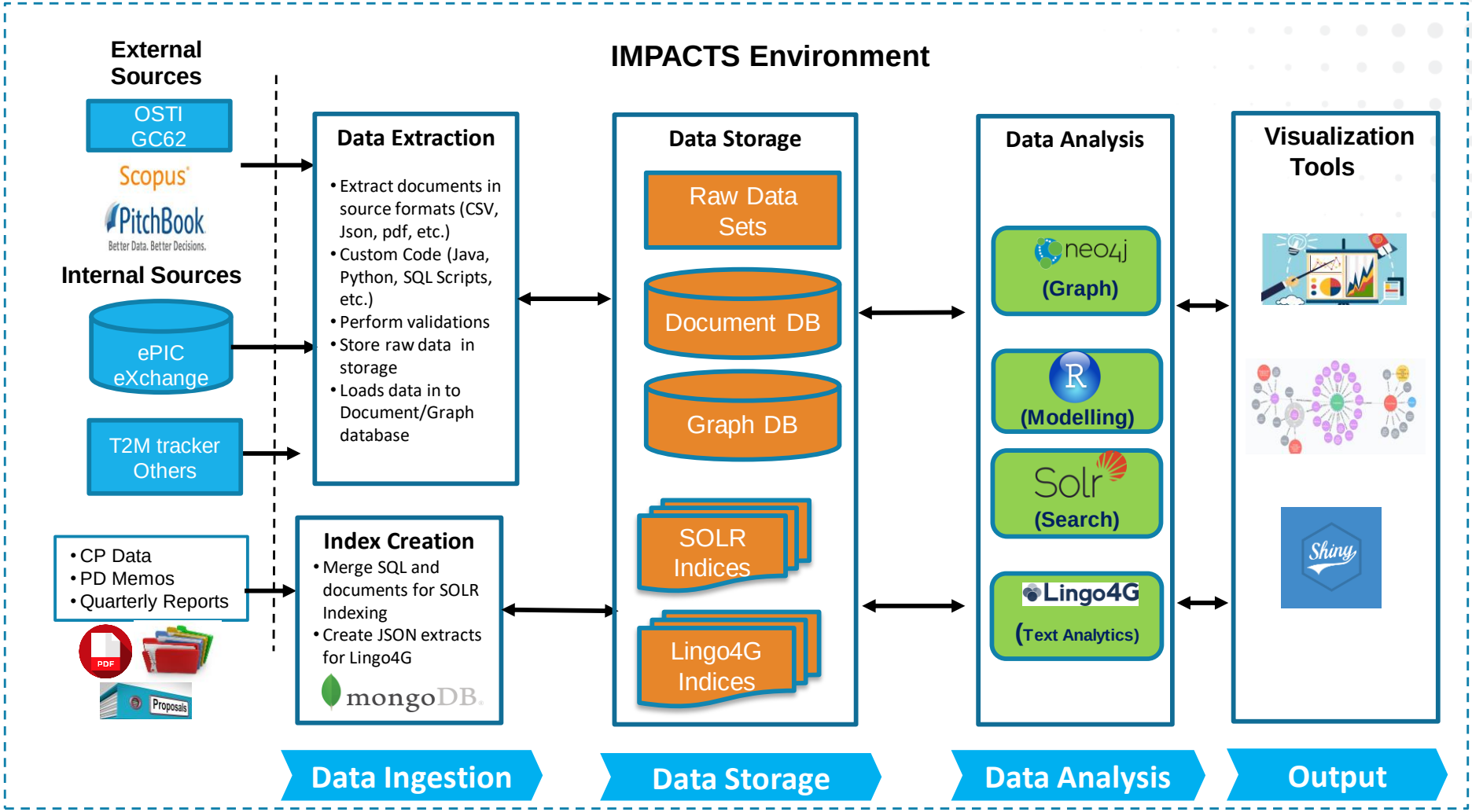
DATA SCIENCE

Write models and view results using statistical analysis and machine learning algorithms

Advanced Research Projects Agency - Energy
U.S. Department of Energy
1000 Independence Ave SW - Washington DC 20585

Questions or comments?
Contact Ann Xu: ann.xu@hq.doe.gov

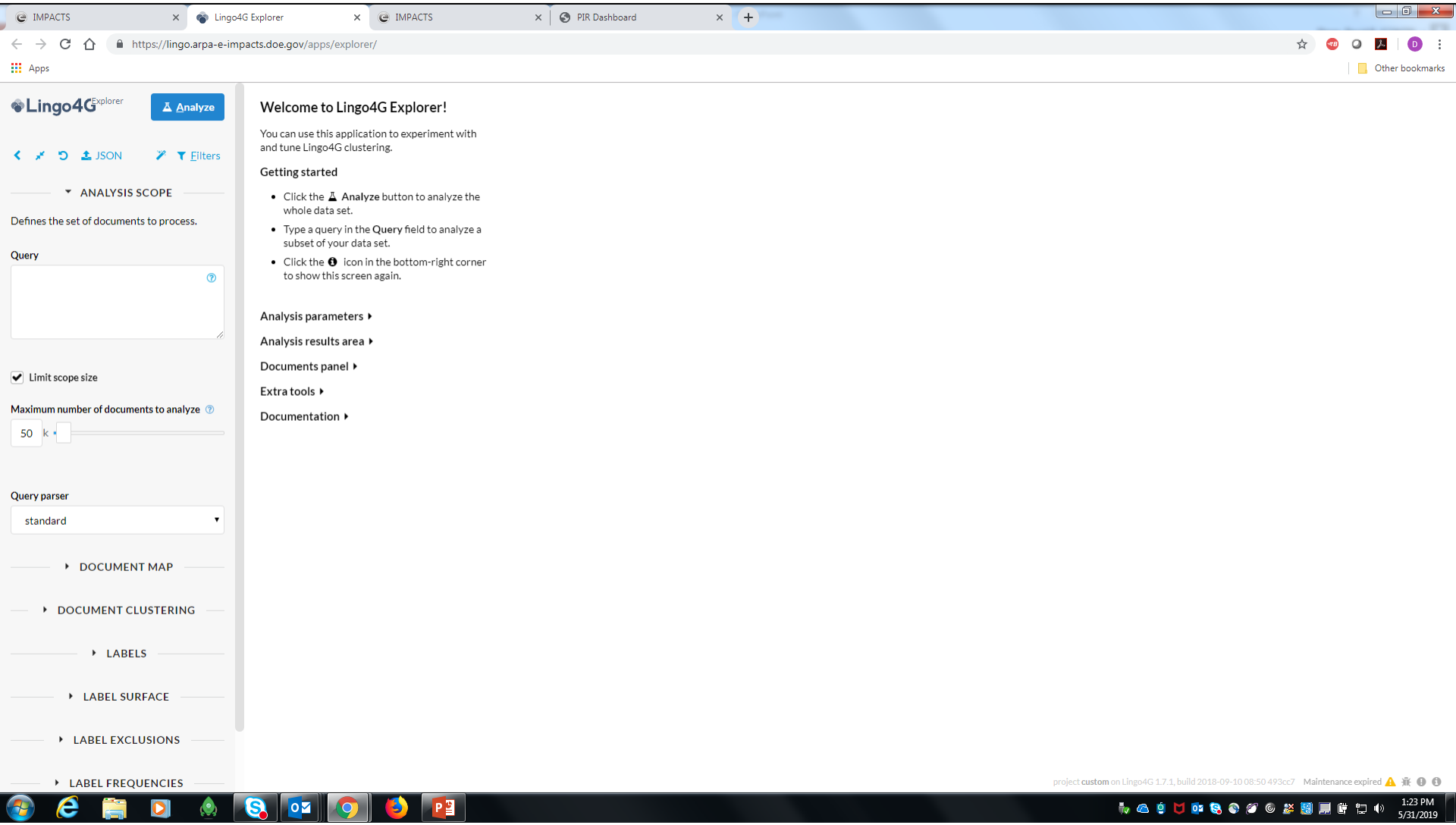
Big data puts rigor in impact assessment



Before Award

PROPOSAL TOPIC ANALYSIS

Text Clustering: Landing page

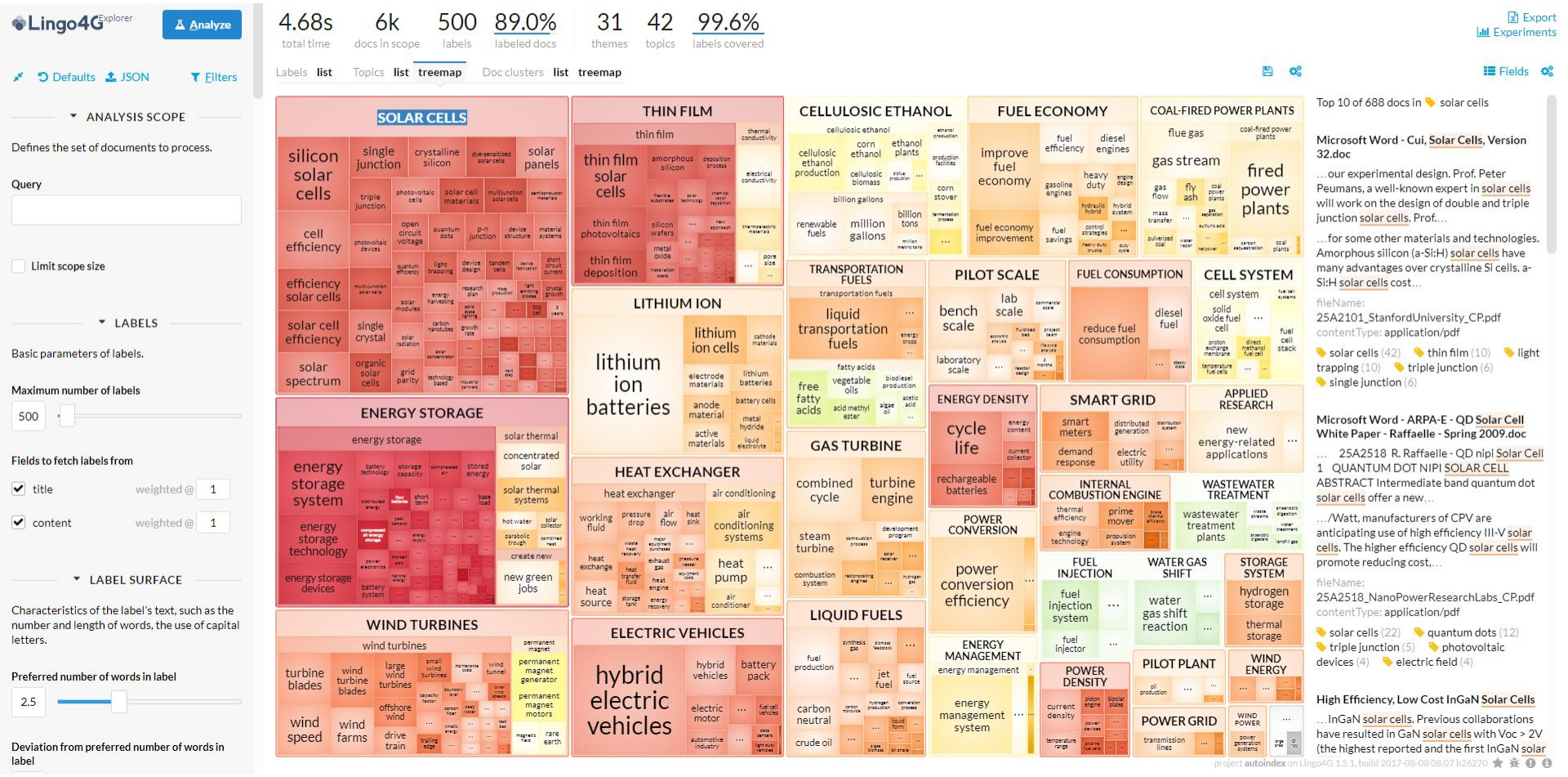


~20k Proposals, ~1 min

OPEN 2018


lead acid batteries
carbon fiber
diesel engine
combustion
engine
energy storage
compressed air
wave energy
energy density
electrode
anode
fuel cells
electrochemical
million tons
capture
electric field
pore size
waste heat
alloys
generator
motor
wind turbine
power systems
microbial fuel cells
hydrogen
reaction
natural gas
liquid fuels
jet fuel
polymer
growth
wastewater treatment
composite
rated power
thin film
solar cells
PV cells
solar energy
thermal energy
heat transfer
fluid flow
cooling
air flow
thermal comfort
management system
power electronics
nuclear power plants
transmission lines
reliability
laser
optical
network
climate change
sensors
remote sensing
machine learning

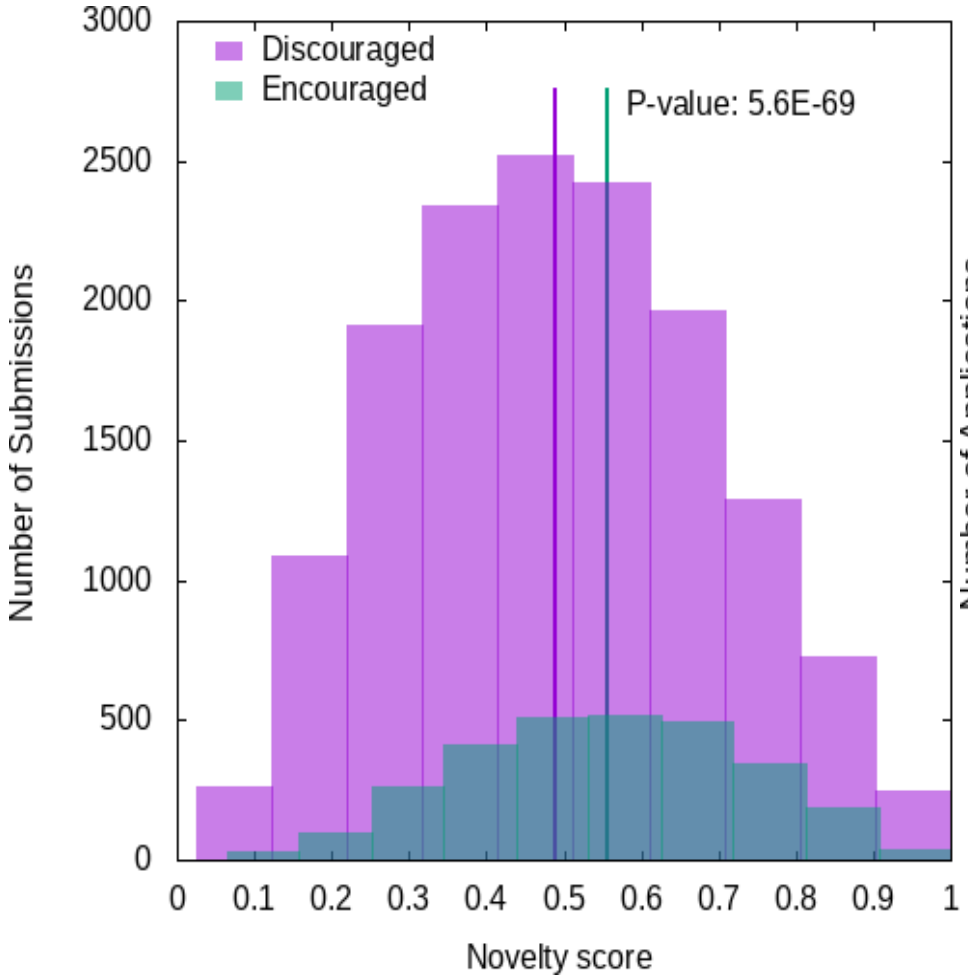
OPEN 2009 Concept Papers

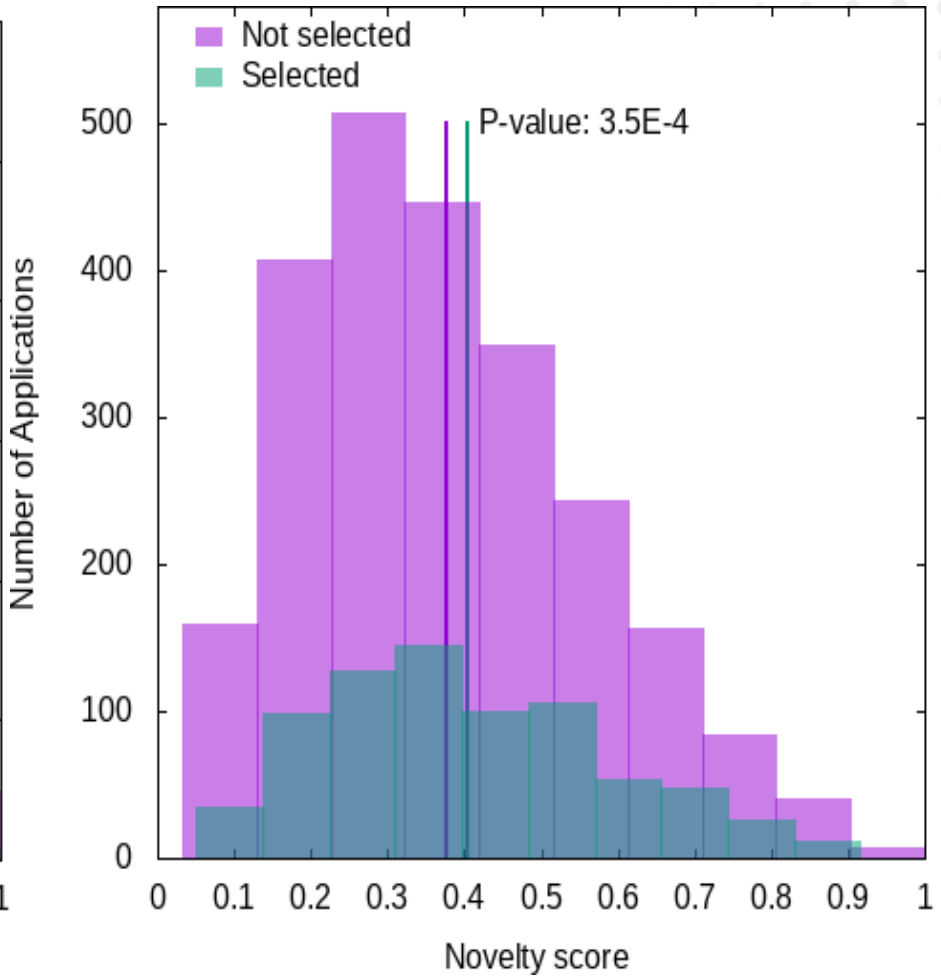


OPEN 2018 Concept Papers

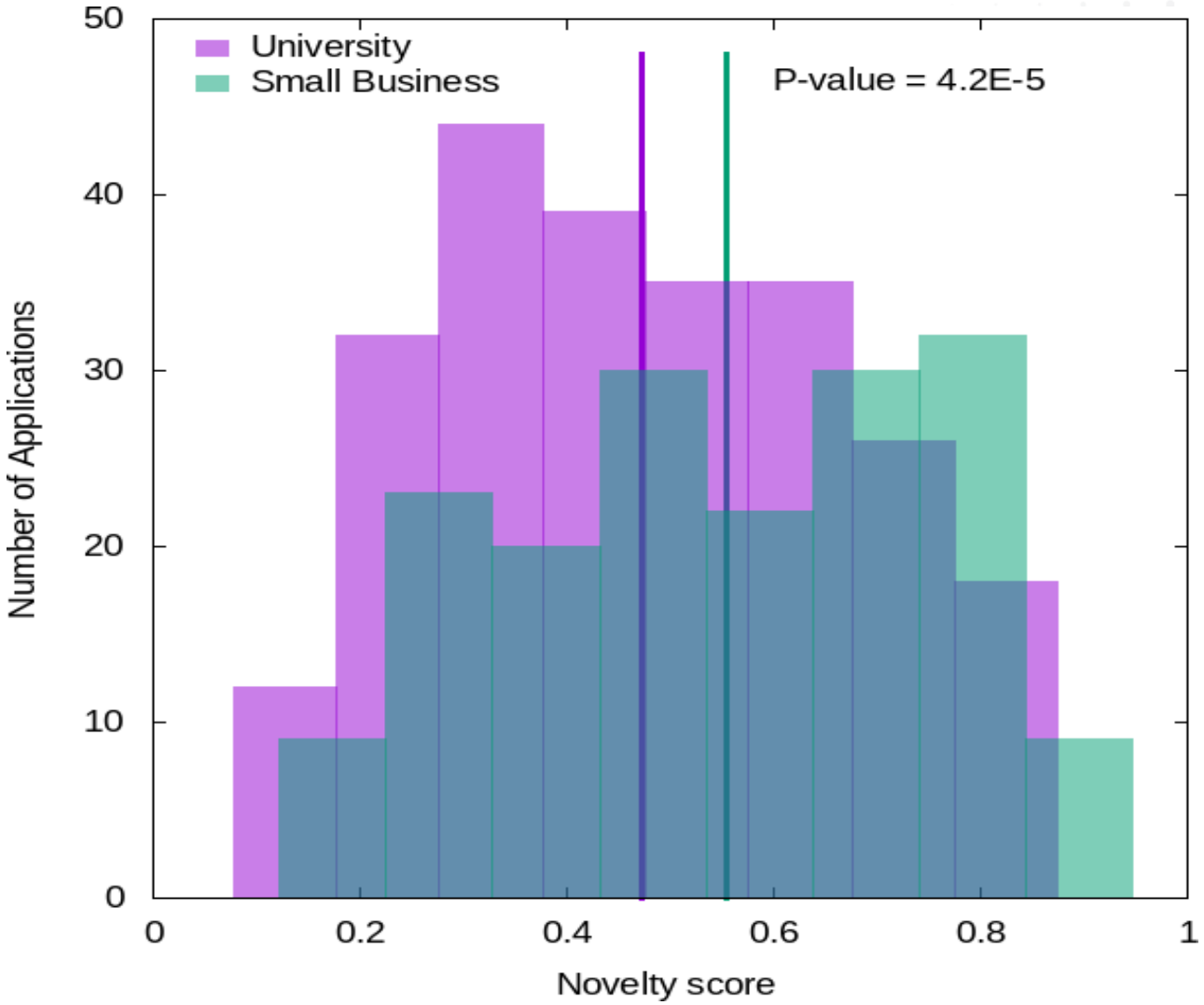
Novelties of ARPA-E concept papers and full apps

More novel CPs are encouraged and more novels FAs are selected





Selected full applications from small businesses are more novel than selected applications from universities



Leveraging External Publication Networks and Author Assessment for New Programs

- Created “Whitespace Analysis” Capability to evaluate new program designs
- A natural language processing based co-citation network is the foundation of searches
- Measure article influence using Betweenness Centrality
- Authors are measured (and able to be ranked by):
  - Aggregate centrality – Measures the influence and information flow from one article to another
  - H-index- Represents the output and influence of an author
  - Output - Average number of citations per year within a range
  - “Rockstar” -Represents an author’s “trajectory” and growth over the past 6 years

FOA Whitespace Analysis: Flexible CCS

Case study:

- Searched publication and author networks to validate that there is not a significant body of literature in the technical areas around a “Flexible CCS” topic area
  - Also identified key SMEs in this field that might be worth engaging in program design
- Searched Pitchbook for relevant “comps” based on a known industry partners to expand knowledge of current commercial space
- Key Benefits:
  - Direct man hr savings: 2 days of SETA labor for this particular case
  - Enhanced search coverage: Significant increase in publication and author coverage using analytical approach
    - 77% increase in relevant authors found
    - 5x increase in publications identified
    - 3x in commercial entity identification
  - IP Search functionalities are still nascent
- Note: this process will not replace the current baseline approach… just enhance coverage and replace some manual labor with machine search capabilities

NLP accelerates knowledge digestion

- Ability to ingest, search, and categorize millions of documents
- Identify trends and themes
- Improve cross-agency coordination and reduce overlap

During Award

INVESTOR MATCH & PROGRAM EVALUATION

Network Analysis: Graph UI Landing Page

IMPACTS | IMPACTS | PIR Dashboard

https://apps.arpa-e-impacts.doe.gov/graphui/

Apps | Other bookmarks

arpa·e

Home | Getting Started | About

Graph UI

Explore the Graph Database

Use Cases | Create Query

Programs under ARPA-E Category:

Select ARPA-E Category

Create Graph

Projects Under Program:

Select Program

Create Graph

Project's associated Organizations and Investors:

Select Investment Type

Select Program

Create Graph

Top Investors in Program Space:

Select Program

Number of Investors...

Create Table

Nodes | Edges

Welcome!

Search ARPA-E Data and find hidden connections using IMPACTS custom Graph UI!

What is a graph?

Information is presented in a form of a graph in which **nodes** are connected by **edges**.

Nodes represent entities like People, Programs or Organizations. Nodes that represent the same entity type have the same color.

Edges represent relationships between entities or events that involve both entities. When an Organization entity is the Prime to an Project entity, the graph contains both entities as nodes of different color and the Prime relationship is represented by an edge between the nodes.

Different types of nodes and relationships will have unique sets of properties that define them.

Where to start?

Test out any of our 4 common use cases on on the sidebar to the right.

- Get all Programs under an ARPA-E Category
- Get all Projects under a Program
- Get all Organizations and Investors associated with a Project
- Get the Top Investors in a Program Space

The results of a different Use Case queries are independent of each other. You can perform one Use Case query at a time.

How to create a custom query?

How to filter a graph?

How to search a graph with keywords?

What do the buttons on the canvas do?

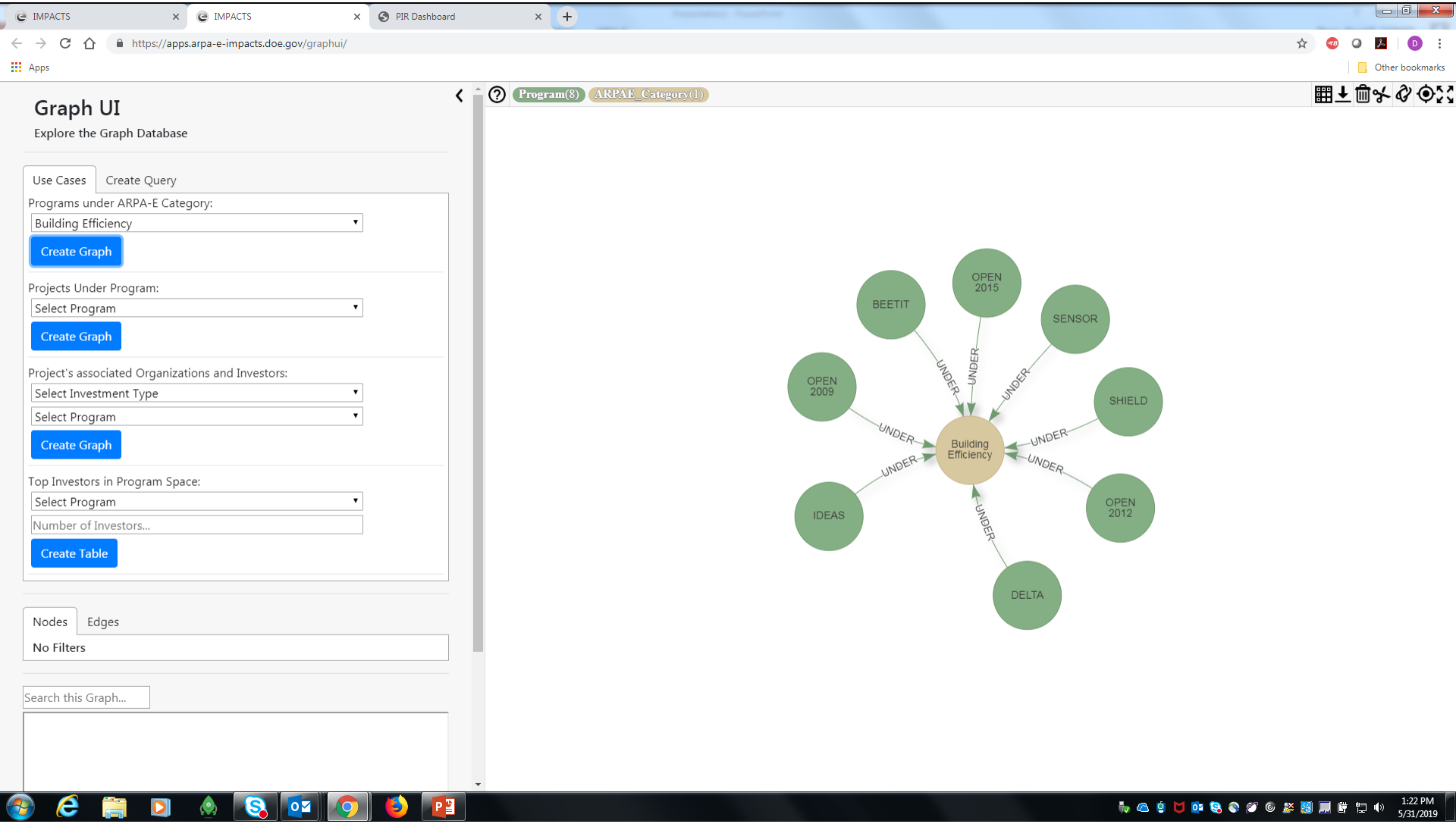
Advanced Research Projects Agency - Energy
U.S. Department of Energy
1000 Independence Ave SW - Washington DC 20585

Questions or comments?
Contact: ARPA-E-IMPACTS-Support@hq.doe.gov

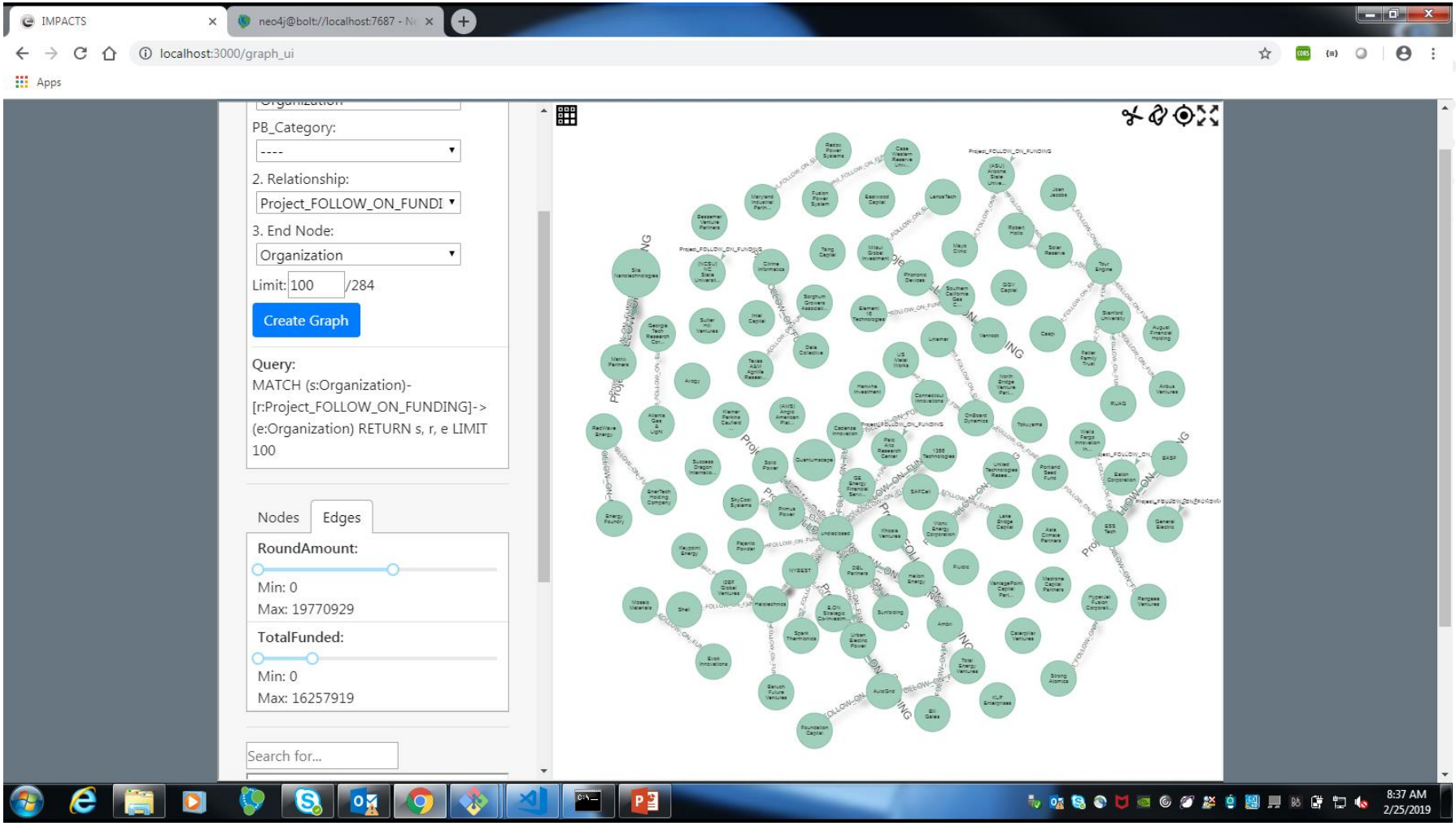
NOTICE TO USERS: This is a Federal computer system and is the property of the United States Government. This system is for authorized use only. All users (authorized

1:13 PM
5/31/2019

Sample Use Case: Programs in Industry Cat

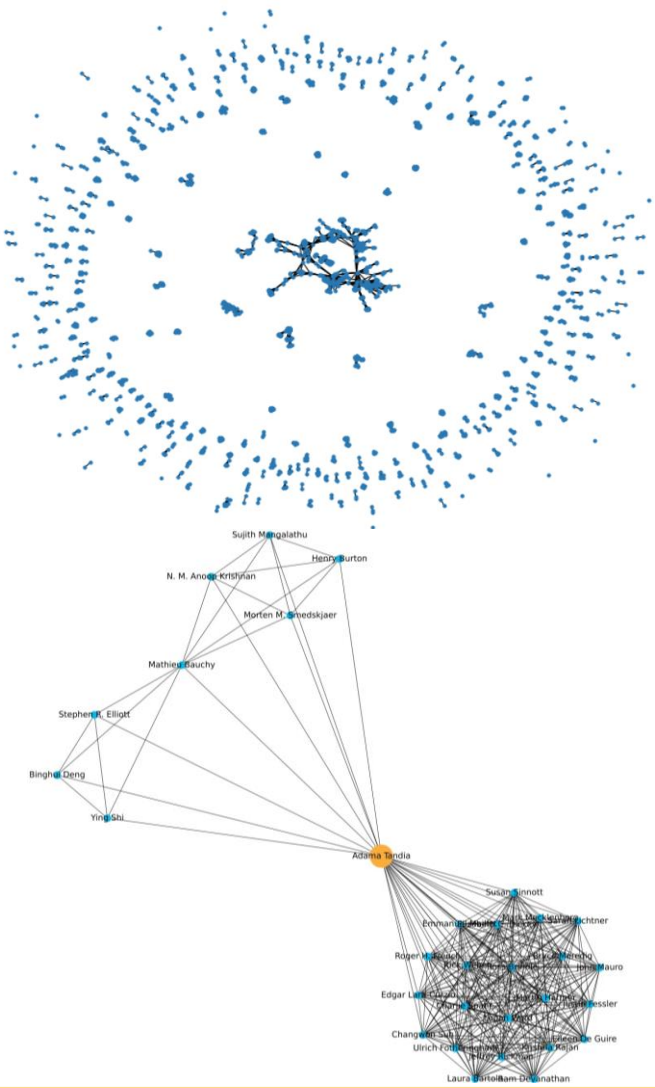


Use Case: Investment/Project Tracker



Case Study: Targeted analysis of research landscape for DIFFERENTIATE

- Motivation
  - Determine if a program concept conceived of in 2018 is still cutting edge in 2020
- Methodology
  - Query Scopus and arXiv with provided keywords (physics informed machine learning) to get articles
  - Extract author lists, journal titles, keywords, abstracts
  - Combined with IMPACTS white space analysis to get related articles and authors using most highly connected authors' papers as seeds
- Results
  - 705 articles with 2035 authors retrieved from queries
  - Top authors included workshop speaker
  - **Identified previously unknown authors that are heavily connected and influential in networks to work with ARPA-E Teams**



After Award

PORTFOLIO ANALYSIS

Advanced Data Reporting: Landing Page

IMPACTS | PIR Dashboard

https://dashboard.arpa-e-impacts.doe.gov

Apps | Other bookmarks

PIR Dashboard

- About PIR Dashboard
- **Choose PIR Version:** PIR_2019-04-15_FINAL.csv
- **Analysis by:** All (Web Only Projects)
- **Selection:** All (Web Only Projects)
- **Select Interactive Asset:**
  - Sankey Diagrams
  - Project Outcomes
  - Outcome Drivers
  - PIR Data Table
- Download PIR Data
- **Select Static Asset:**
  - Analytics Gallery (Pending)

About the Portfolio Impact Review Dashboard

The **Portfolio Impact Review (PIR)** is a snapshot of ARPA-E funded projects from the perspective of project management, project output, and industry activity – according to project technology focus. From the project management perspective, data elements that are included in the PIR describe various ARPA-E resources that have been assigned to a project. These include, but not limited to, management personnel, funding actions and project performance. In term of project output, the PIR includes data elements that describes the various artifacts produced by a project team. These artifacts range from publications, to subject inventions and number of patents issued, to follow-on-funding and the formation of new companies. The PIR also captures industry activity by extracting investment related data to the relevant technology areas. The information that these data element presents include the amount of funding being invested in a given technology area, the size of these investments and the volatility associated with the investment, in a given technology area.

The PIR Dashboard sits on a reservoir of data that is the PIR. This dashboard is intended to enable user, throughout ARPA-E, to quickly gain insights from APRA-E's portfolio of projects either or within targeted sub-areas, further enabling (pro)active project management and drive impact. Currently, the dashboard allows users to explore the PIR by Program, Technology Type, Project Size, and Project Status. Whether exploring the PIR in its entirety or by a subset of project, this dashboard is intended to enable users to extract insights and derive ultimate actions to remain in lockstep with ARPA-E's (pro)active project management to derive the most value out of the agency's investments.

How to Use:

Step 1: Select the version of the PIR you would like to explore. By default, the most recent version of the PIR is pre-loaded.

Step 2: Choose the type of analysis you would like to conduct from the "Analysis by" drop-down and then the "Selection" drop-down. Upon making a selection, all interactive assets of the dashboard will be updated accordingly. By default, all dashboard assets are rendered based on the data from the most recent PIR.

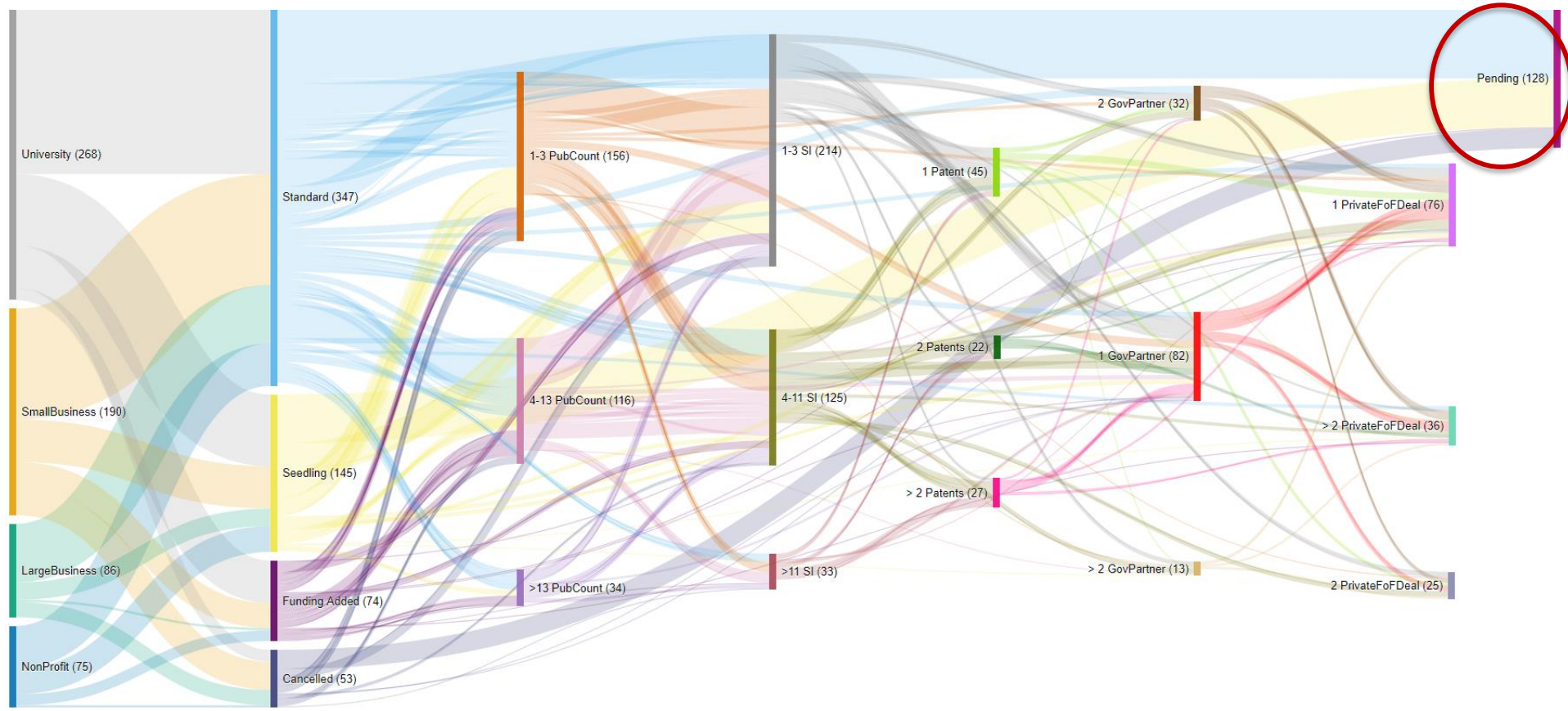
Step 3: Select an interactive dashboard asset to view the output from the drop-down selections. The interactive dashboard assets can be accessed from the "Sankey Diagram", "Project Outcomes", "Outcome Drivers" and "PIR Data Table" tabs.

Step 4 (Optional): Click the "Download PIR Data" button to download either the full PIR dataset or a subset of the PIR based on current analysis. A .xlsx file will be downloaded to your download folder, with the filename beginning with 'PIR_AnalysisExport_'. This file will also contain a data dictionary tab.

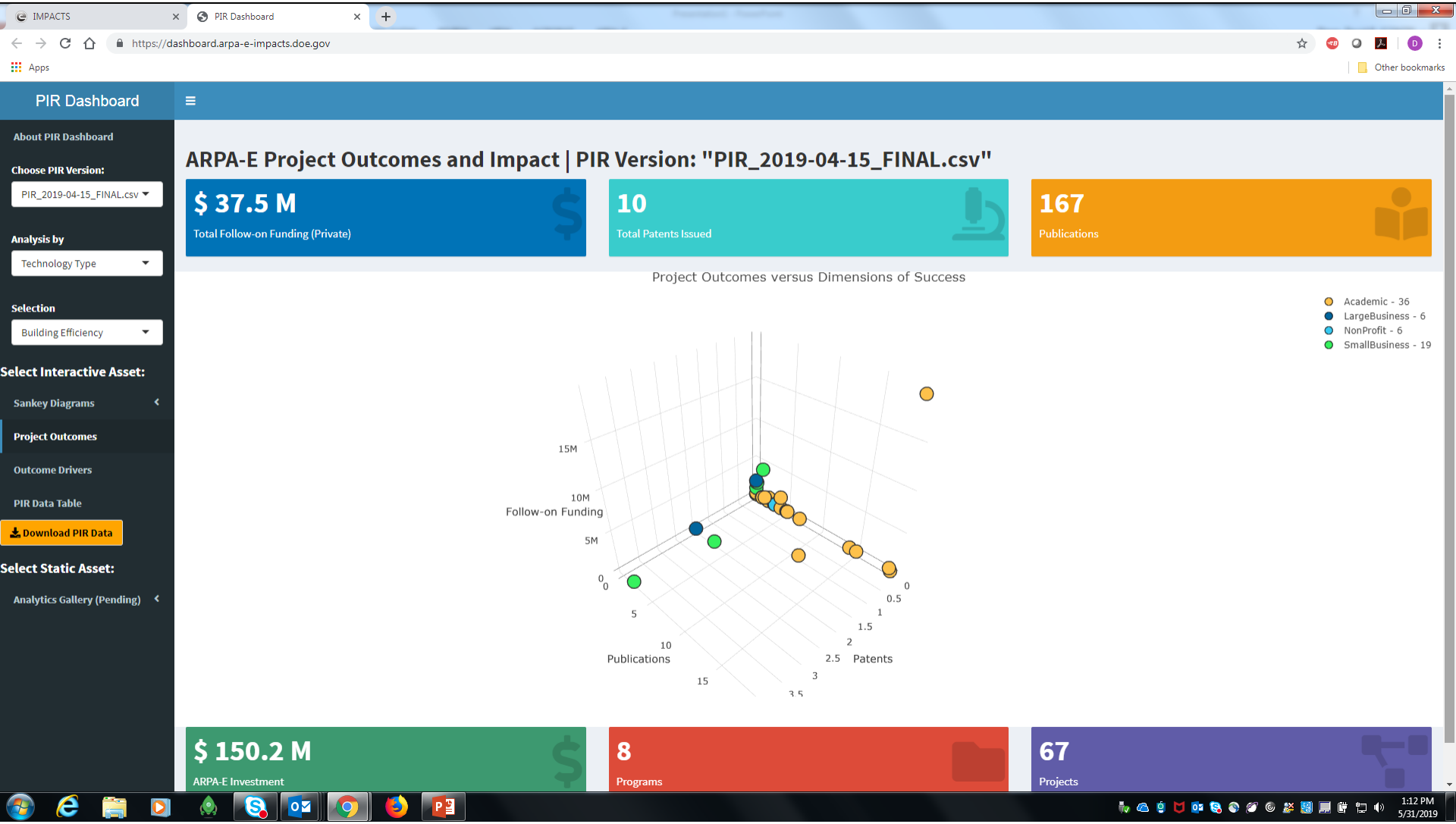
Note: The Analytics Gallery showcases a subset of analyses, conducted by the IMPACTS team that were enabled by the PIR. This is a living Gallery and will be updated as additional PIR use cases get submitted by ARPA-E to the IMPACTS team.

1:11 PM 5/31/2019

Bird's Eye View – How much risk is appropriate for ARPA-E?

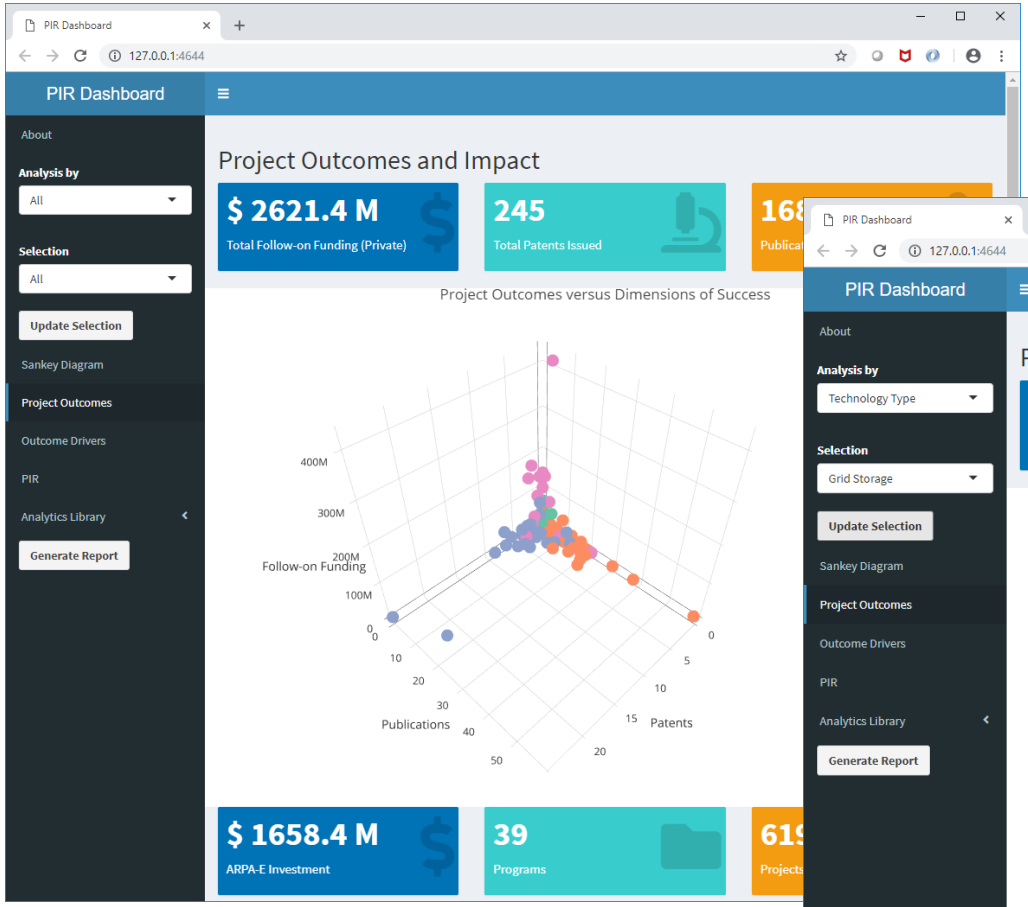


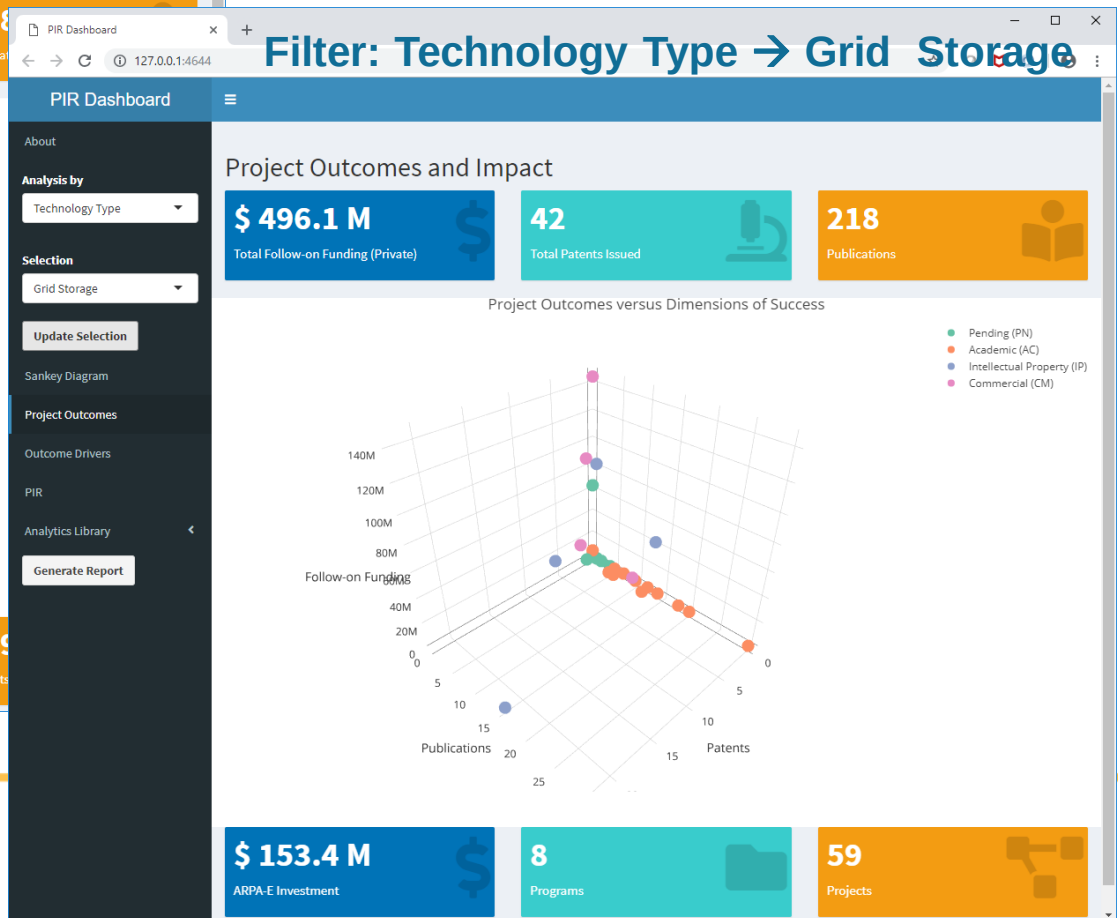
Advanced Data Reporting: Clusters



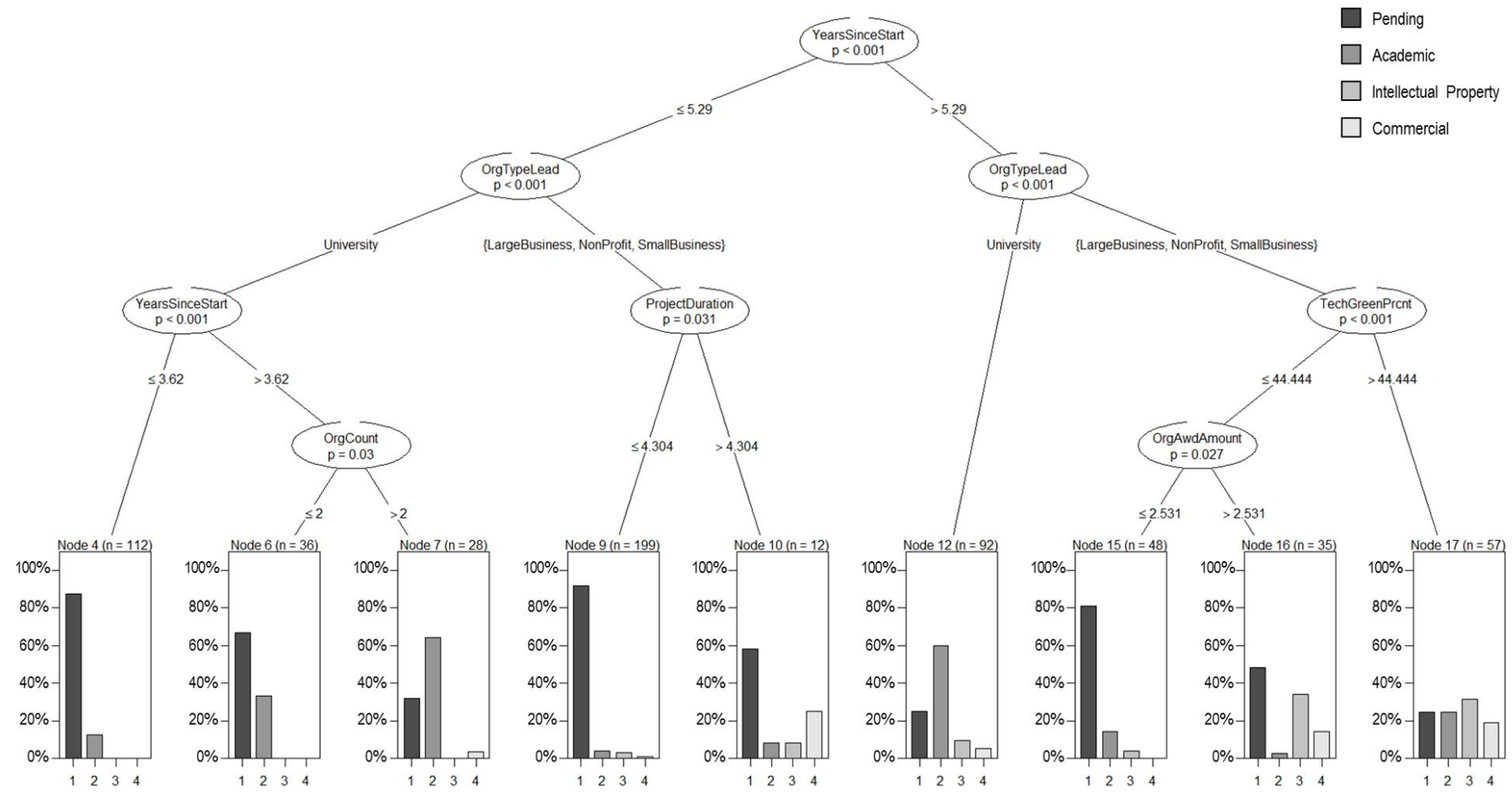
Advanced Data Reporting: Filters

Enabling analyses by Program, Technology Type and Project Status

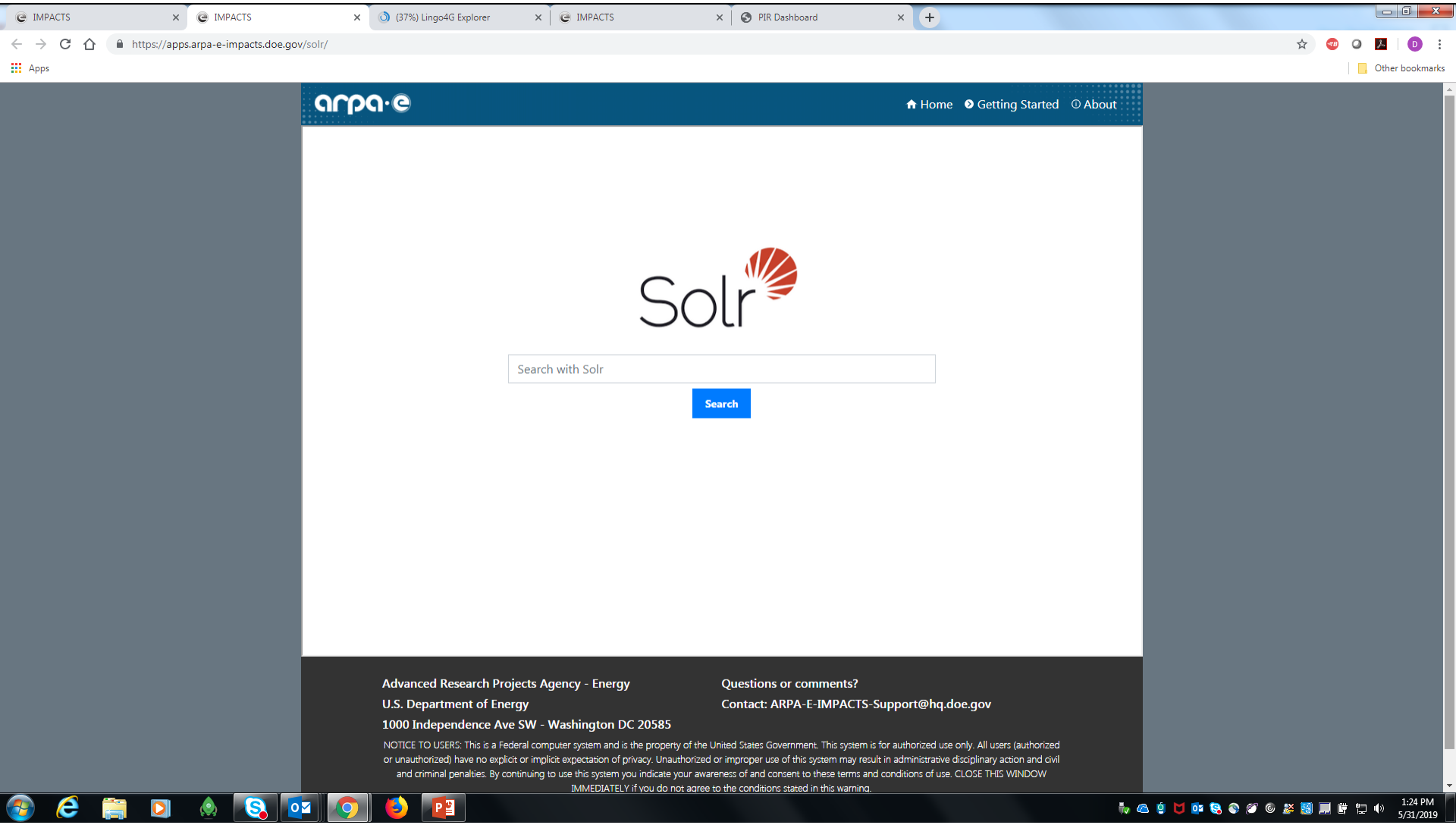




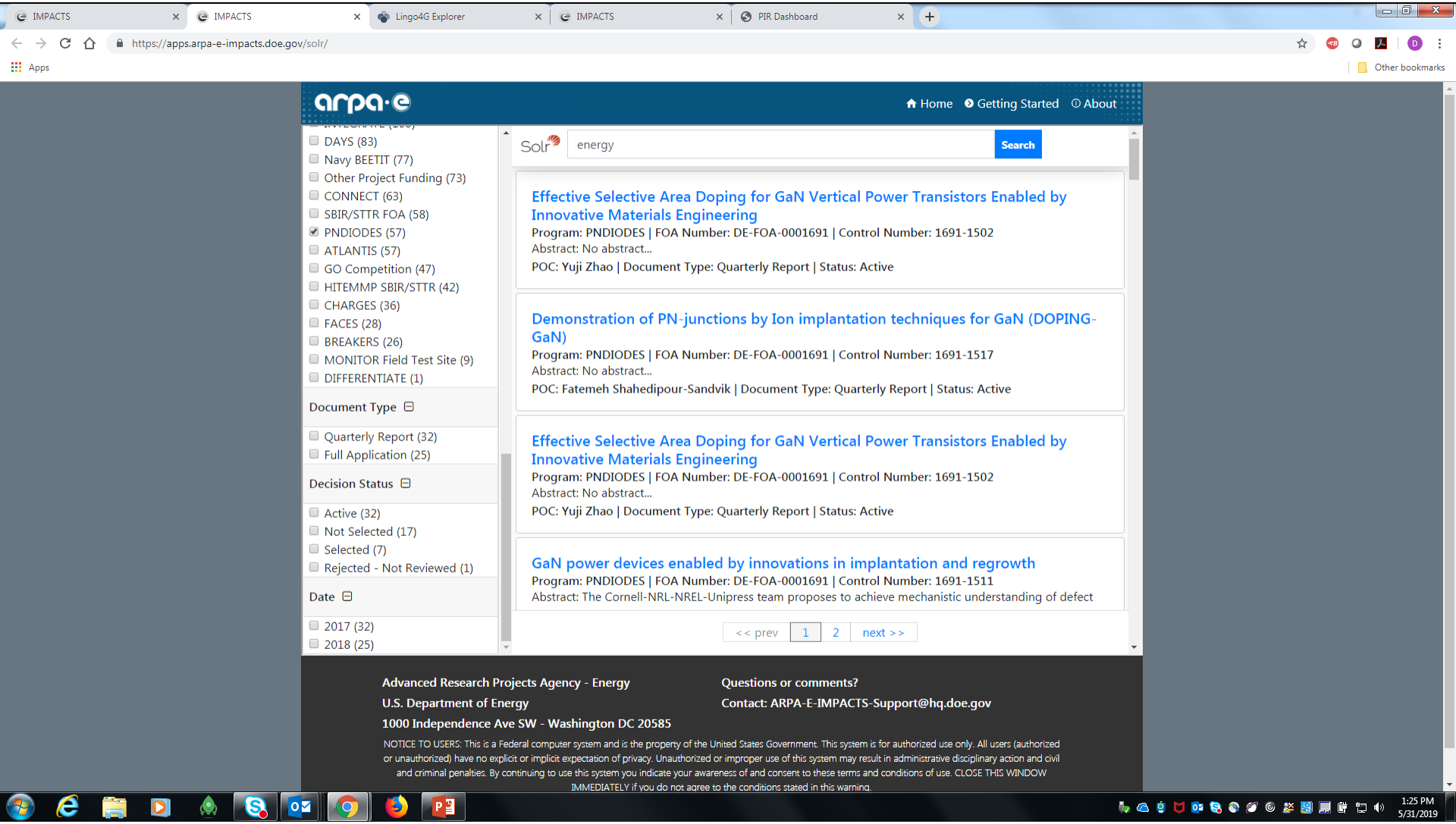
Classification Tree – How do you tell what factors matter to R&D success?



Full Text Search: Landing Page



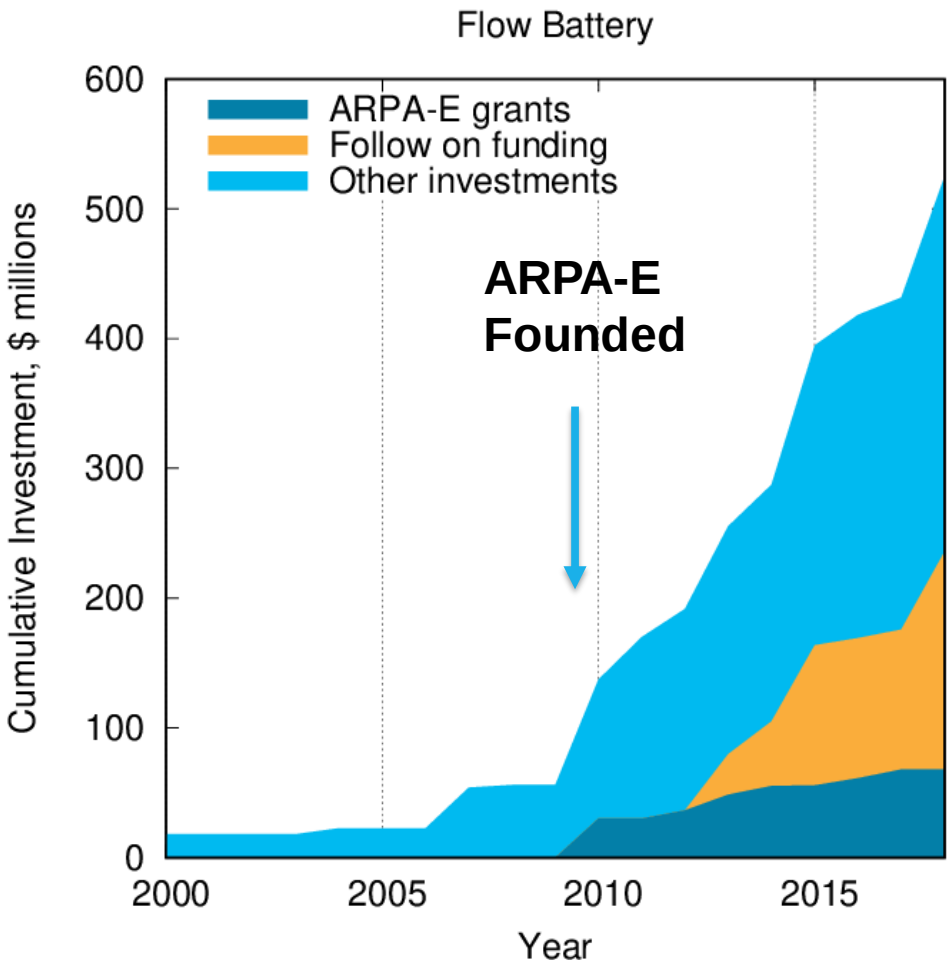
Full Text Search: Results



Before, During and After Program Launch

CASE STUDIES

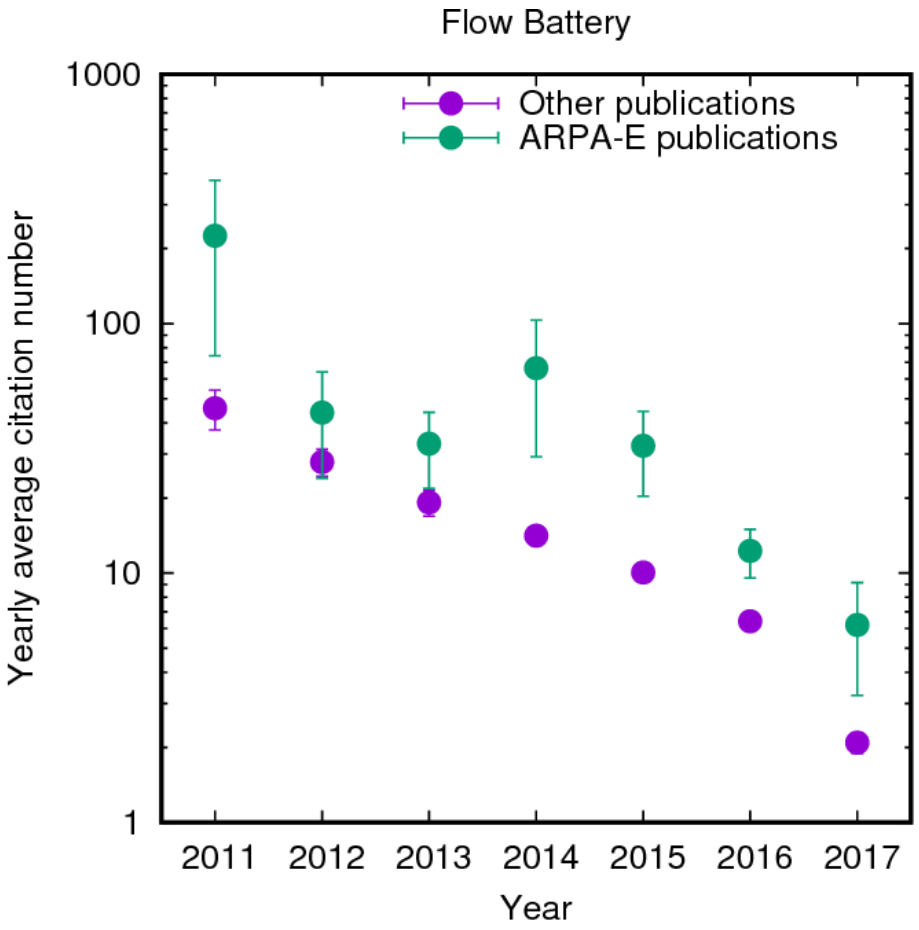
ARPA-E grants spurred industry funding



"Other Investments" = Pitchbook deals not associated with ARPA-E teams in this space

ARPA-E grants were a substantial fraction of investment in the Flow Battery technology early on. Later, as external funding levels increased, ARPA-E dialed its grants down.

ARPA-E publications are highly cited



Direct evidence that ARPA-E publications are more highly cited

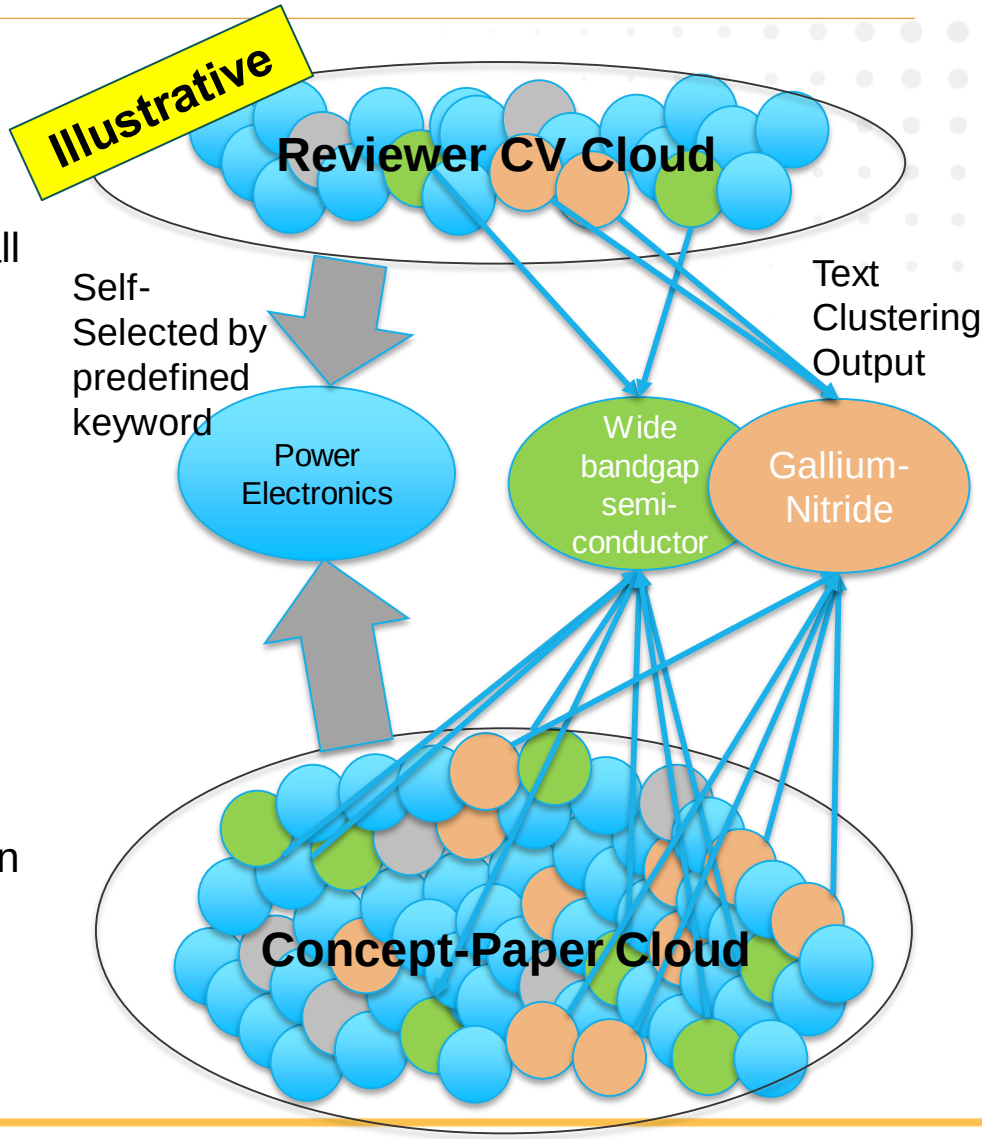
The error bars are the Standard Error of the Mean for all papers published in a particular year

Case Study: Analyzing Project Portfolio Data to Identify Intellectual Property

- The farther away from the period of performance, the lower ARPA-E's visibility into the project outcomes are
- Overall data quality is poor especially on products and licenses
- Methodology:
  - Utilized two machine learning algorithm methodologies on project data, to help identify SI under-reporting to develop a list of potential under reporters
  - Conducted outreach with awardees to request updates in iEdison and acquired external patent databases
- Results:
  - The hit rate is the number of projects for which there is agreement between the prediction and reality divided by the number of projects for which a prediction attempted
  - The percentage of results that were predicted to be yes OR Hit Rate for each model:
    - Regression Based Model = 63%
    - Random Forest Model = 53%
    - GC-62 Manual Methodology=61%
  - The regression-based modeling approach had an equal predictive power to the GC-62 manual approach, although the two data sets only overlapped slightly

Case Study: Leveraging IMPACTS for Concept Paper Reviewer Match

- Methodology
  - Received 175 concept papers (CPs)
  - Developed document clustering map using full text clustering software for all CPs
  - Identified semantic similarity of text groupings that aligned to technology areas of interest
- Results
  - Analyzed an additional 580 reviewers CVs in EXCHANGE database and recommended an additional 100 reviewers beyond initial target list of 50
  - Applied best fit to ensure no more than 8 CPs per reviewer
  - Approximately 50% successful match rate on first try
  - Reduced staff labor hrs by 1 week



Questions?

https://arpa-e.energy.gov