

# 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 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**

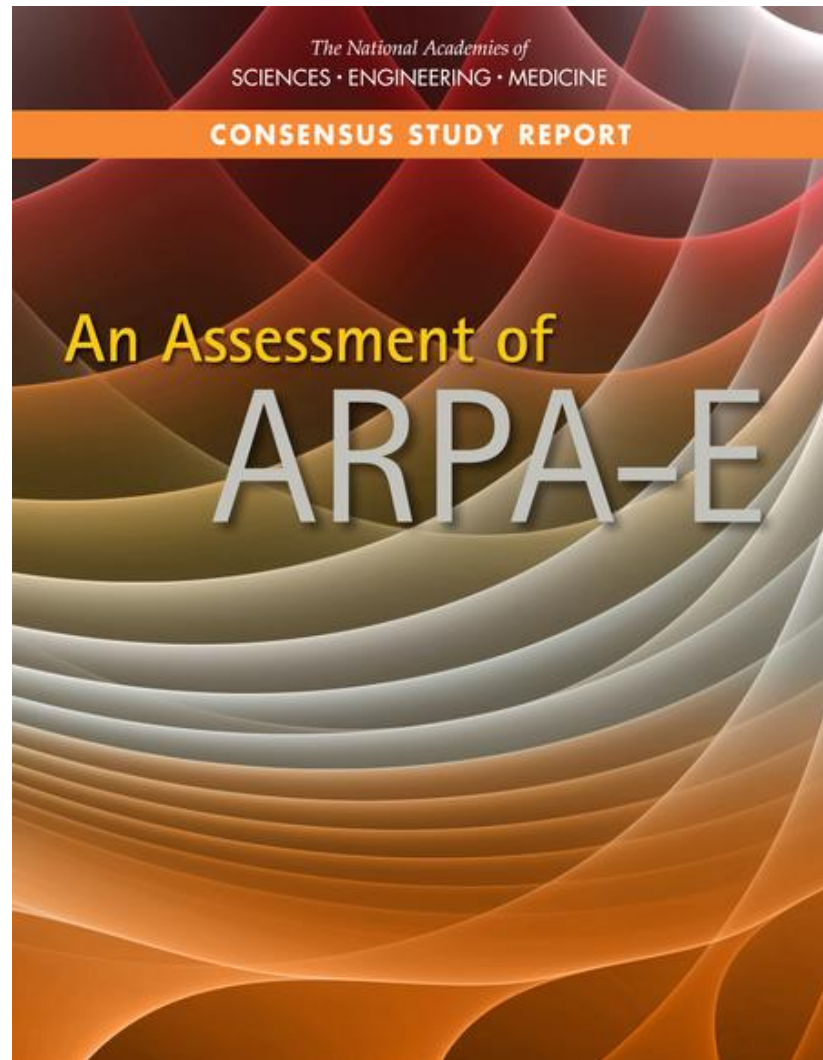
issued by U.S.  
Patent and  
Trademark Office



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:

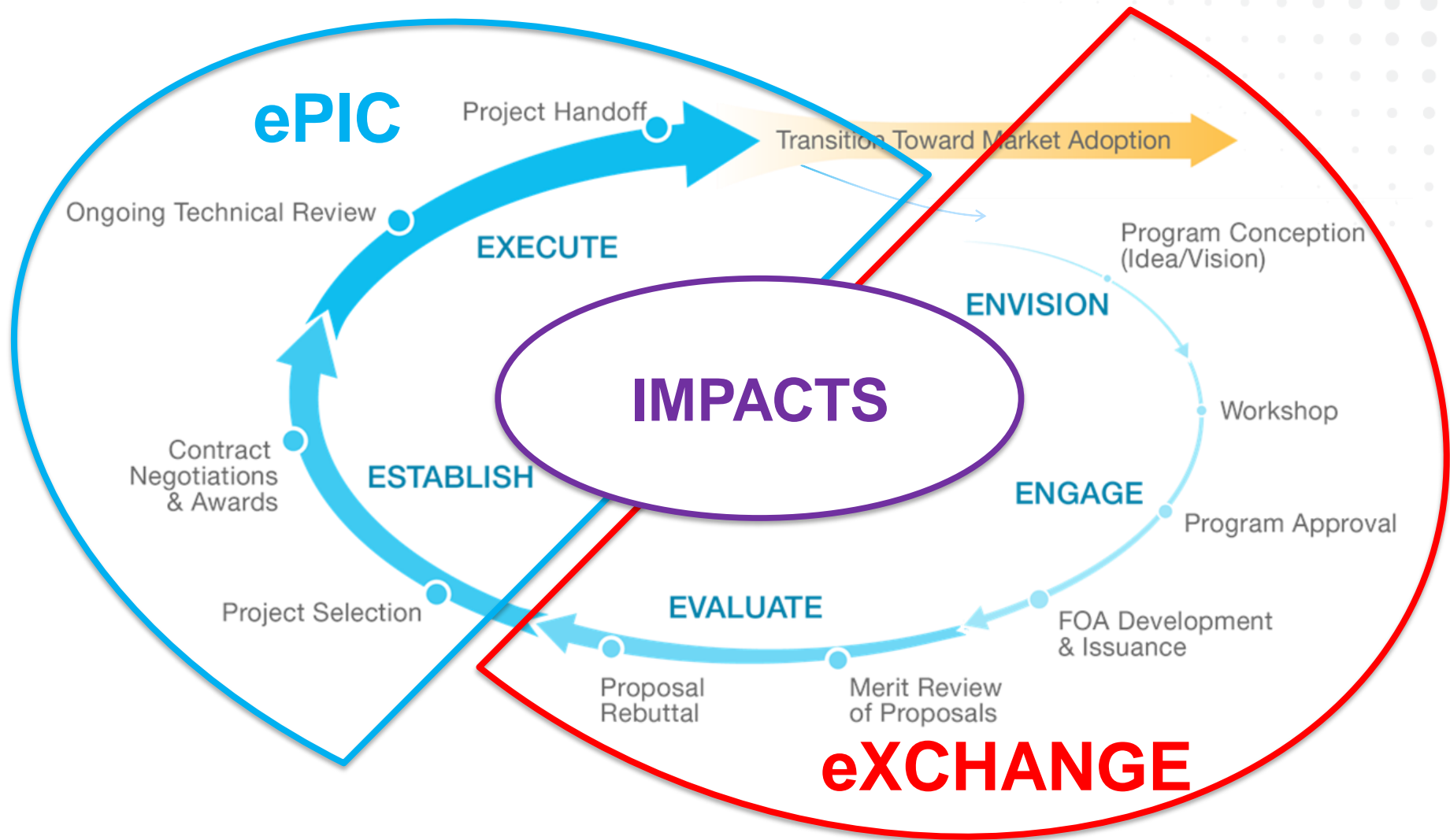
**“The ARPA-E director and program directors should develop and implement a framework for measuring and assessing the agency’s impact in achieving its mission and goals.”**

# 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

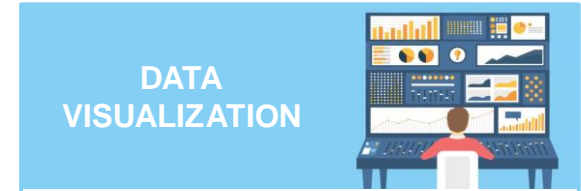


- Automatic data Ingestion
  - Python
- NoSQL storage
  - Graph database
  - Document database
  - Structured & un-structured
  - Scalable for future needs




## DATA ANALYSIS

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



## DATA VISUALIZATION

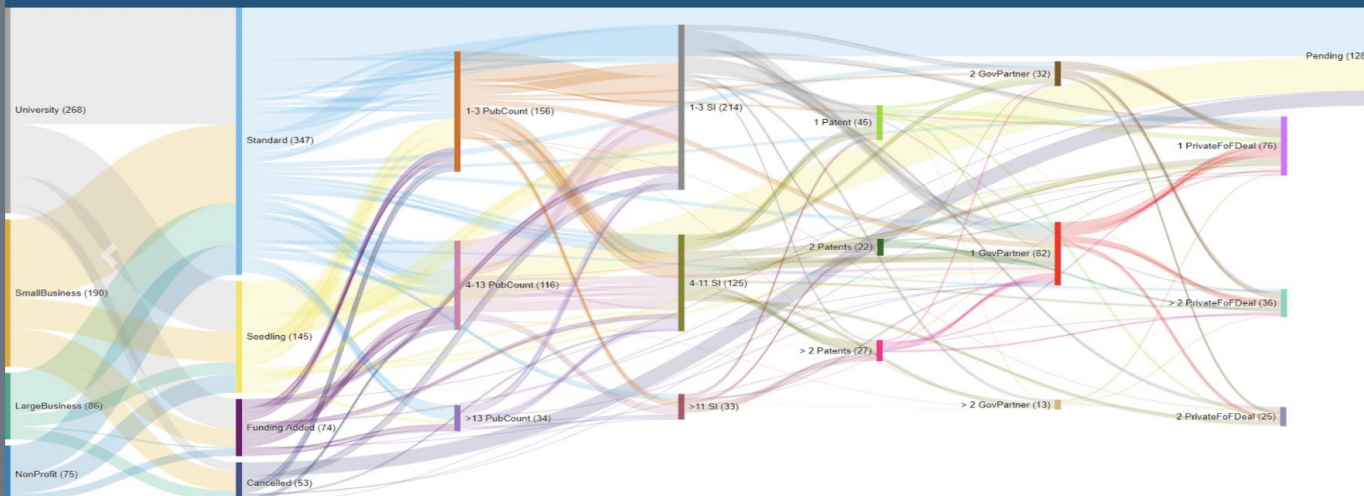
- Discover hidden relationships in highly connected data
- Identify patterns and clusters
- Turbo-charge communications





# 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

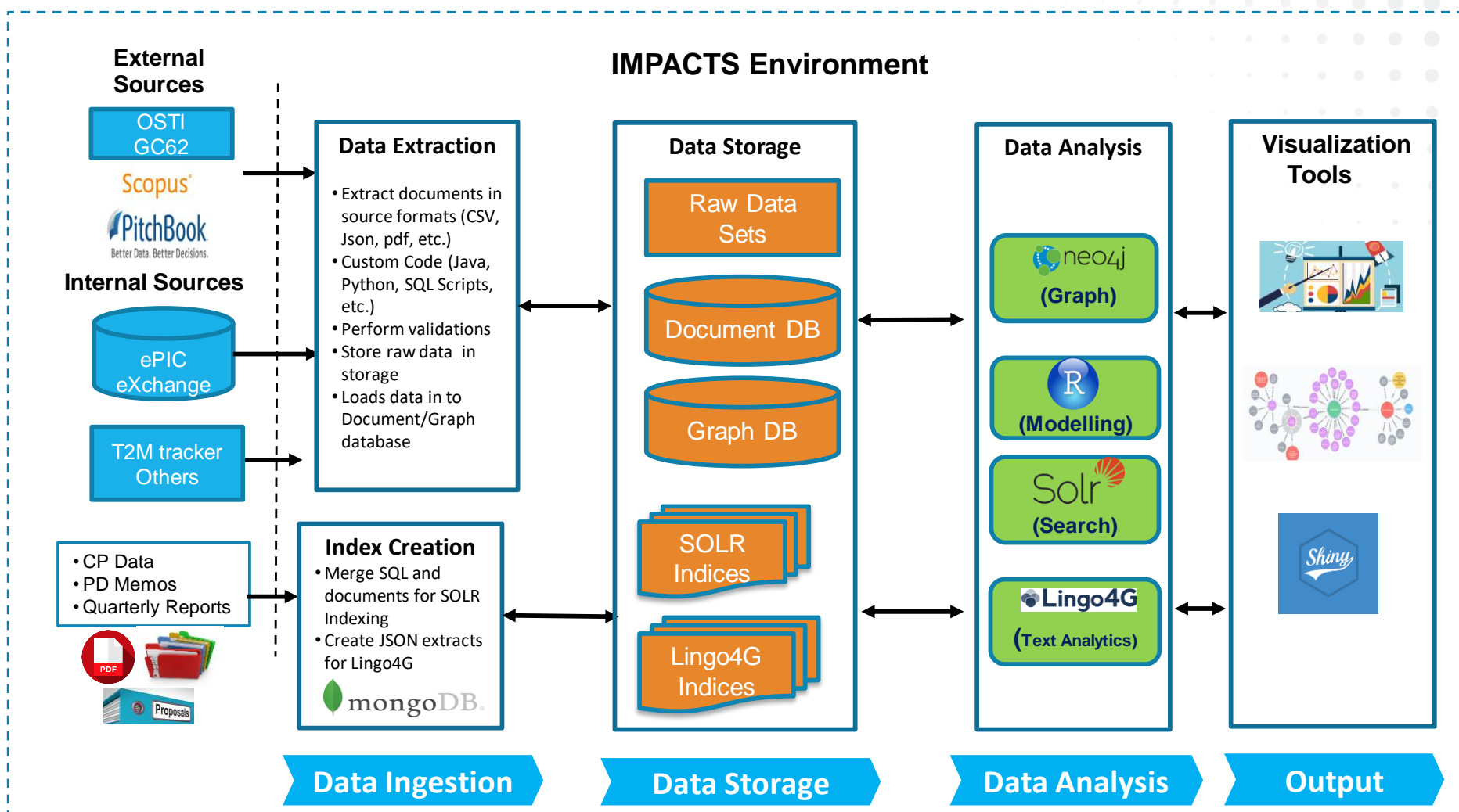


## DATA SCIENCE

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



# Big data puts rigor in impact assessment





Before Award

# PROPOSAL TOPIC ANALYSIS

# Text Clustering: Landing page

The screenshot shows a web browser with multiple tabs. The active tab is 'Lingo4G Explorer' at the URL <https://lingo.arpa-e-impacts.doe.gov/apps/explorer/>. The browser's address bar shows the URL, and the page title is 'Lingo4G Explorer'. The application interface is divided into a left sidebar and a main content area.

**Left Sidebar:**

- Lingo4G Explorer** (with an 'Analyze' button)
- Navigation icons: back, forward, home, search, JSON, Filters
- ANALYSIS SCOPE** (expanded)
  - Defines the set of documents to process.
  - Query** (text input field)
  - ☒ Limit scope size
  - Maximum number of documents to analyze (50 k)
  - Query parser** (dropdown menu showing 'standard')
- DOCUMENT MAP**
- DOCUMENT CLUSTERING** (expanded)
  - LABELS**
  - LABEL SURFACE**
  - LABEL EXCLUSIONS**
  - LABEL FREQUENCIES**

**Main Content Area:**

- Welcome to Lingo4G Explorer!**

You can use this application to experiment with and tune Lingo4G clustering.
- Getting started**
  - Click the **Analyze** button to analyze the whole data set.
  - Type a query in the **Query** field to analyze a subset of your data set.
  - Click the **?** icon in the bottom-right corner to show this screen again.
- Analysis parameters** (arrow)
- Analysis results area** (arrow)
- Documents panel** (arrow)
- Extra tools** (arrow)
- Documentation** (arrow)

At the bottom of the browser window, a status bar shows: 'project custom on Lingo4G 1.7.1, build 2018-09-10 08:50 493cc7 Maintenance expired 1:23 PM 5/31/2019'.

# ~20k Proposals, ~1 min





# OPEN 2018



# OPEN 2009 Concept Papers

**Lingo4G** Explorer

[Analyze](#)

Defaults JSON Filters

4.68s total time 6k docs in scope 500 labels 89.0% labeled docs 31 themes 42 topics 99.6% labels covered

Labels list Topics list **treemap** Doc clusters list treemap

Export Experiments

## ANALYSIS SCOPE

Defines the set of documents to process.

Query

☐ Limit scope size

## LABELS

Basic parameters of labels.

Maximum number of labels

500

Fields to fetch labels from

☒ title weighted @ 1  
☒ content weighted @ 1

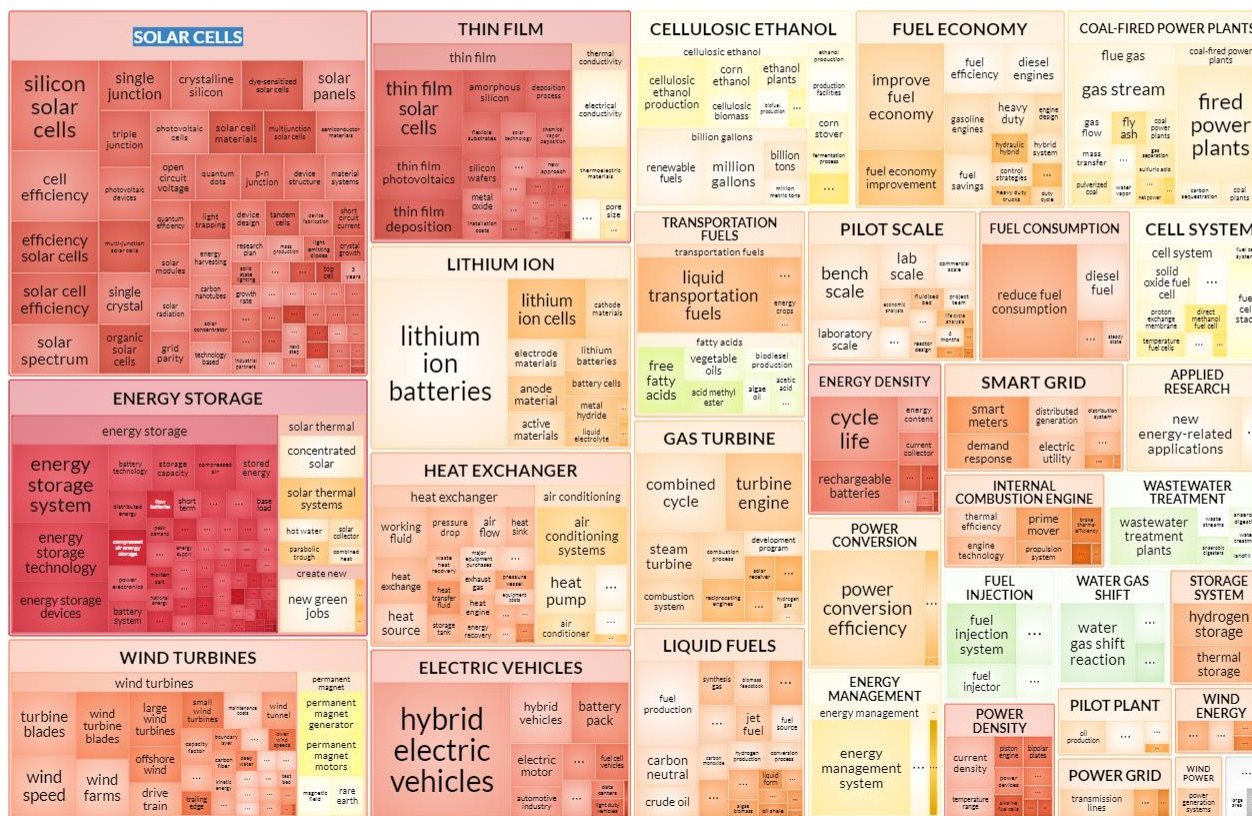
## LABEL SURFACE

Characteristics of the label's text, such as the number and length of words, the use of capital letters.

Preferred number of words in label

2.5

Deviation from preferred number of words in label



Top 10 of 688 docs in solar cells

Microsoft Word - Cui, Solar Cells, Version 32.doc

... our experimental design. Prof. Peter Peumans, a well-known expert in solar cells will work on the design of double and triple junction solar cells. Prof...

... for some other materials and technologies. Amorphous silicon (a-Si:H) solar cells have many advantages over crystalline Si cells, a-Si:H solar cells cost...

fileName: 25A2101\_StanfordUniversity\_CP.pdf  
contentType: application/pdf

🔍 solar cells (42) 🔍 thin film (10) 🔍 light trapping (10) 🔍 triple junction (6) 🔍 single junction (6)

Microsoft Word - ARPA-E - QD Solar Cell White Paper - Raffaele - Spring 2009.doc

... 25A2518 R. Raffaele - QD nipi Solar Cell 1. QUANTUM DOT NIPi SOLAR CELL ABSTRACT Intermediate band quantum dot solar cells offer a new...

.../Watt, manufacturers of CPV are anticipating use of high efficiency III-V solar cells. The higher efficiency QD solar cells will promote reducing cost...

fileName: 25A2518\_NanoPowerResearchLabs\_CP.pdf  
contentType: application/pdf

🔍 solar cells (22) 🔍 quantum dots (12) 🔍 triple junction (5) 🔍 photovoltaic devices (4) 🔍 electric field (4)

High Efficiency, Low Cost InGaN Solar Cells

... InGaN solar cells. Previous collaborations have resulted in GaN solar cells with Voc > 2V (the highest reported and the first InGaN solar

project autoindex on Lingo4G 1.5.1, build 2017-06-08 08:07 b26270



**arpa·e**  
CHANGING WHAT'S POSSIBLE

Deviation from preferred number of words in label

[illegible]

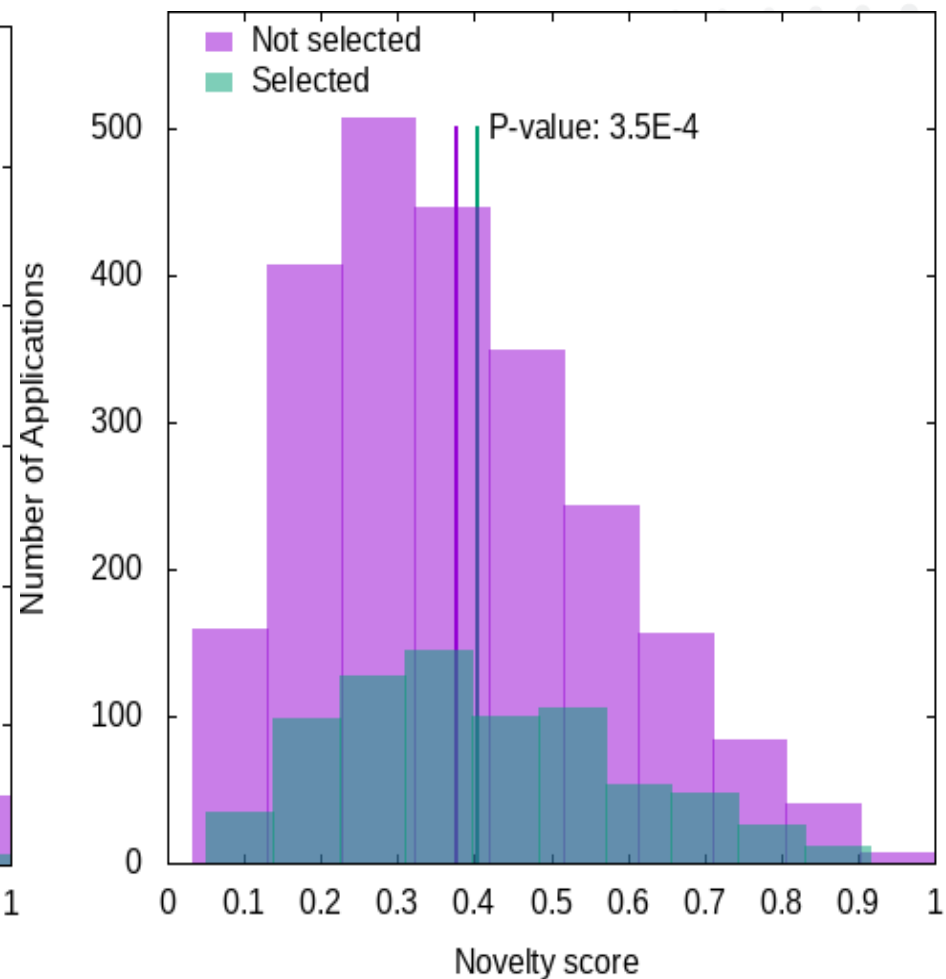
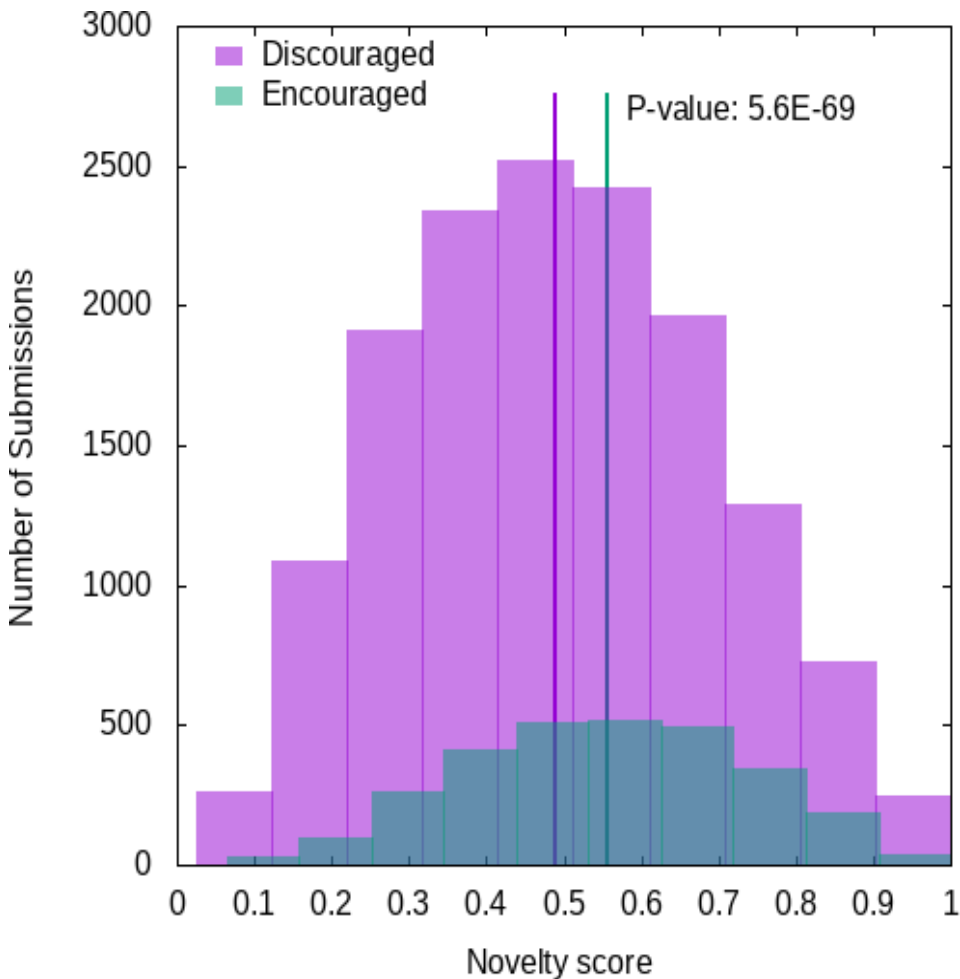
...-09.16 INNOVATIVE TECHNOLOGIES FOR WASTE HEAT POWERED LOW COST WASTE WATER TREATMENT AND REUSE Virginia Tech (Blacksburg, Virginia); Principal Investigator...

project json on Lingo4G 15.1, build 2017-08-08 08:07 b26270

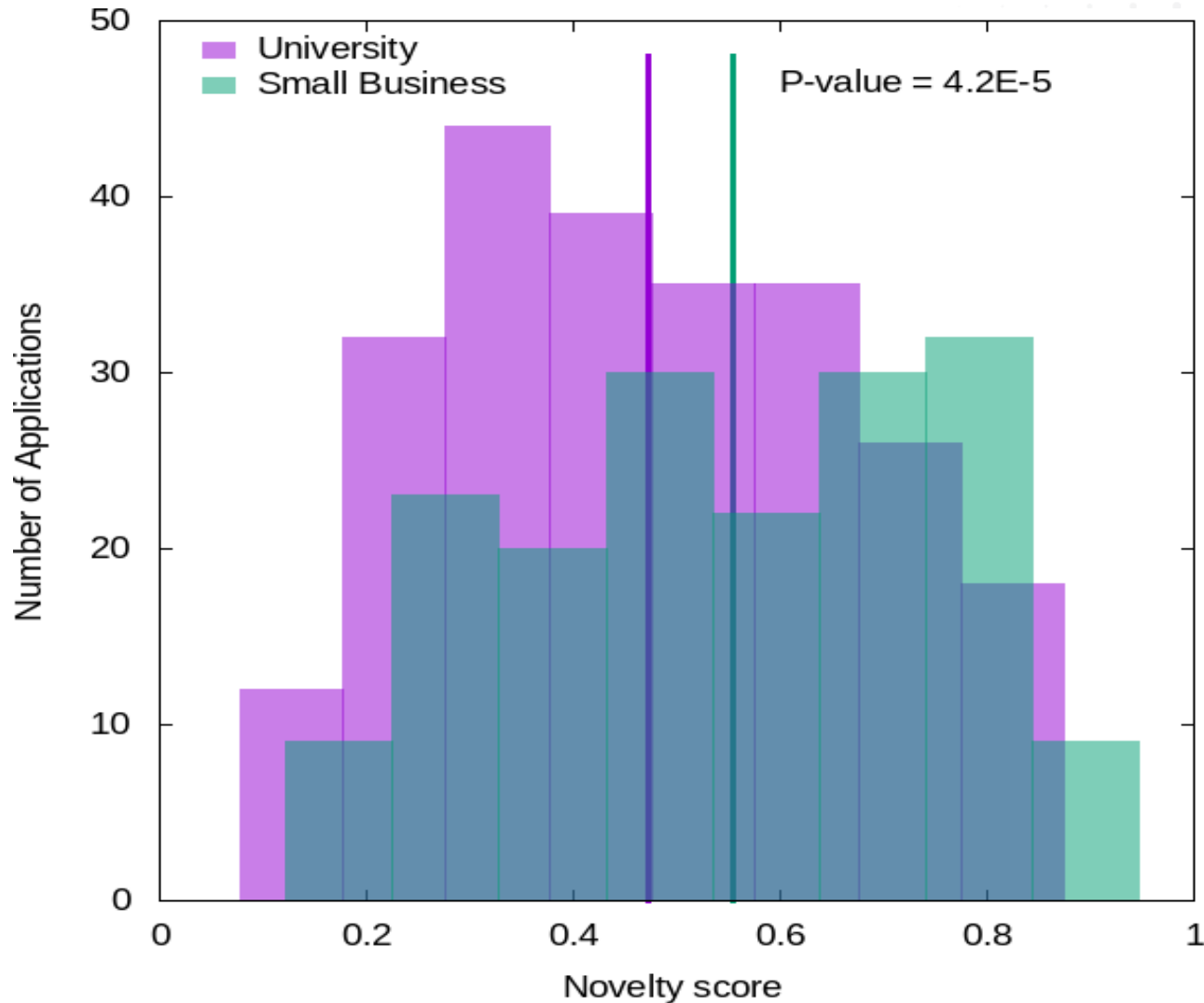


# Novelties of ARPA-E concept papers and full apps

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

The screenshot shows a web browser window with the URL <https://apps.arpa-e-impacts.doe.gov/graphui/>. The page is the Graph UI landing page for the ARPA-E IMPACTS system. The header includes the ARPA-E logo and navigation links for Home, Getting Started, and About. The main content area is divided into two sections: a sidebar on the left for filtering and a main panel on the right for information and use cases.

**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.

**OrgPrimeProject**  
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 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](mailto: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

# Sample Use Case: Programs in Industry Cat

IMPACTS IMPACTS PIR Dashboard

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

Apps Other bookmarks

## Graph UI

Explore the Graph Database

Use Cases Create Query

Programs under ARPA-E Category:

Building Efficiency

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

No Filters

Search this Graph...

Program(8) ARPAE\_Category(1)

```
graph TD; BEETIT -- UNDER --> BE; OPEN2015((OPEN 2015)) -- UNDER --> BE; SENSOR -- UNDER --> BE; SHIELD -- UNDER --> BE; OPEN2012((OPEN 2012)) -- UNDER --> BE; DELTA -- UNDER --> BE; IDEAS -- UNDER --> BE; OPEN2009((OPEN 2009)) -- UNDER --> BE; BE((Building Efficiency))
```

Windows Taskbar: 1:22 PM 5/31/2019



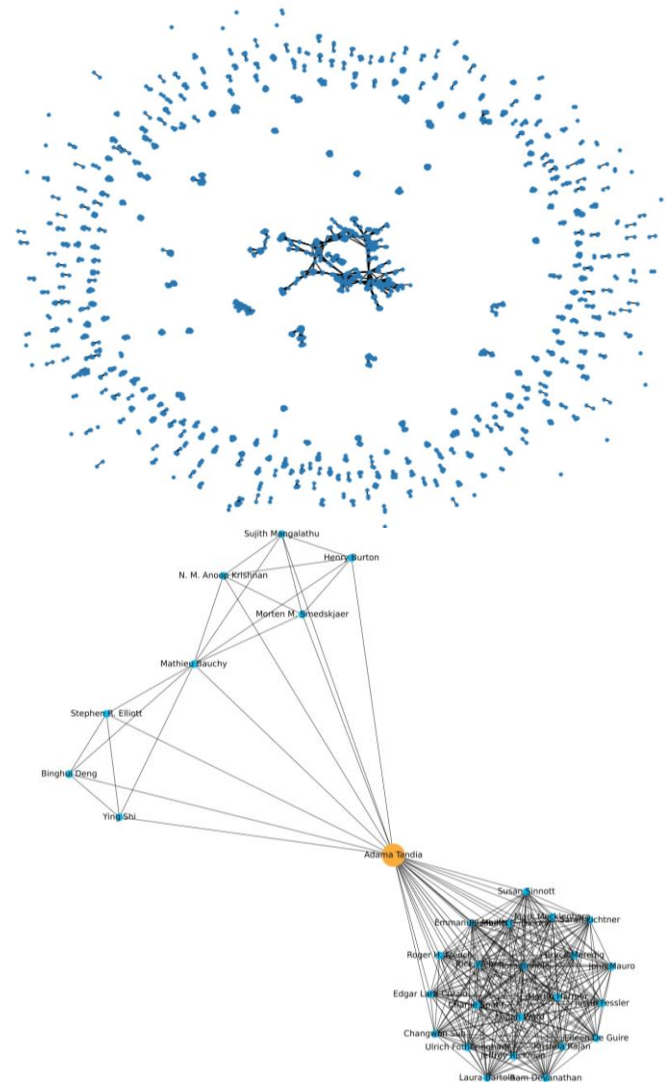


# Case Study: Targeted analysis of research landscape for DIFFERENTIATE

## ► Motivation

## ► Methodology

## ► Results





After Award

# PORTFOLIO ANALYSIS

# Advanced Data Reporting: Landing Page

The screenshot shows a web browser window with two tabs: 'IMPACTS' and 'PIR Dashboard'. The address bar shows the URL 'https://dashboard.arpa-e-impacts.doe.gov'. The browser's address bar and tabs are visible at the top. The page has a blue header with 'PIR Dashboard' and a hamburger menu icon. The left sidebar is dark blue and contains several sections: 'About PIR Dashboard', 'Choose PIR Version:' (with a dropdown menu showing 'PIR\_2019-04-15\_FINAL.csv'), 'Analysis by:' (with a dropdown menu showing 'All (Web Only Projects)'), 'Selection:' (with a dropdown menu showing 'All (Web Only Projects)'), 'Select Interactive Asset:', 'Sankey Diagrams', 'Project Outcomes', 'Outcome Drivers', 'PIR Data Table', and a yellow 'Download PIR Data' button. Below this is 'Select Static Asset:' and 'Analytics Gallery (Pending)'. The main content area has a light blue background and features a large graphic with the 'arpa-e' logo, the text 'Advanced Research Projects Agency • Energy', and 'Portfolio Impact Review'. Below the graphic, there is a 'How to Use:' section with four steps and a note. The bottom of the browser window shows a Windows taskbar with various application icons and a system clock showing 1:11 PM on 5/31/2019.

IMPACTS x PIR Dashboard x +

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.

arpa-e

Advanced Research Projects Agency • Energy

Portfolio Impact Review

### 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.

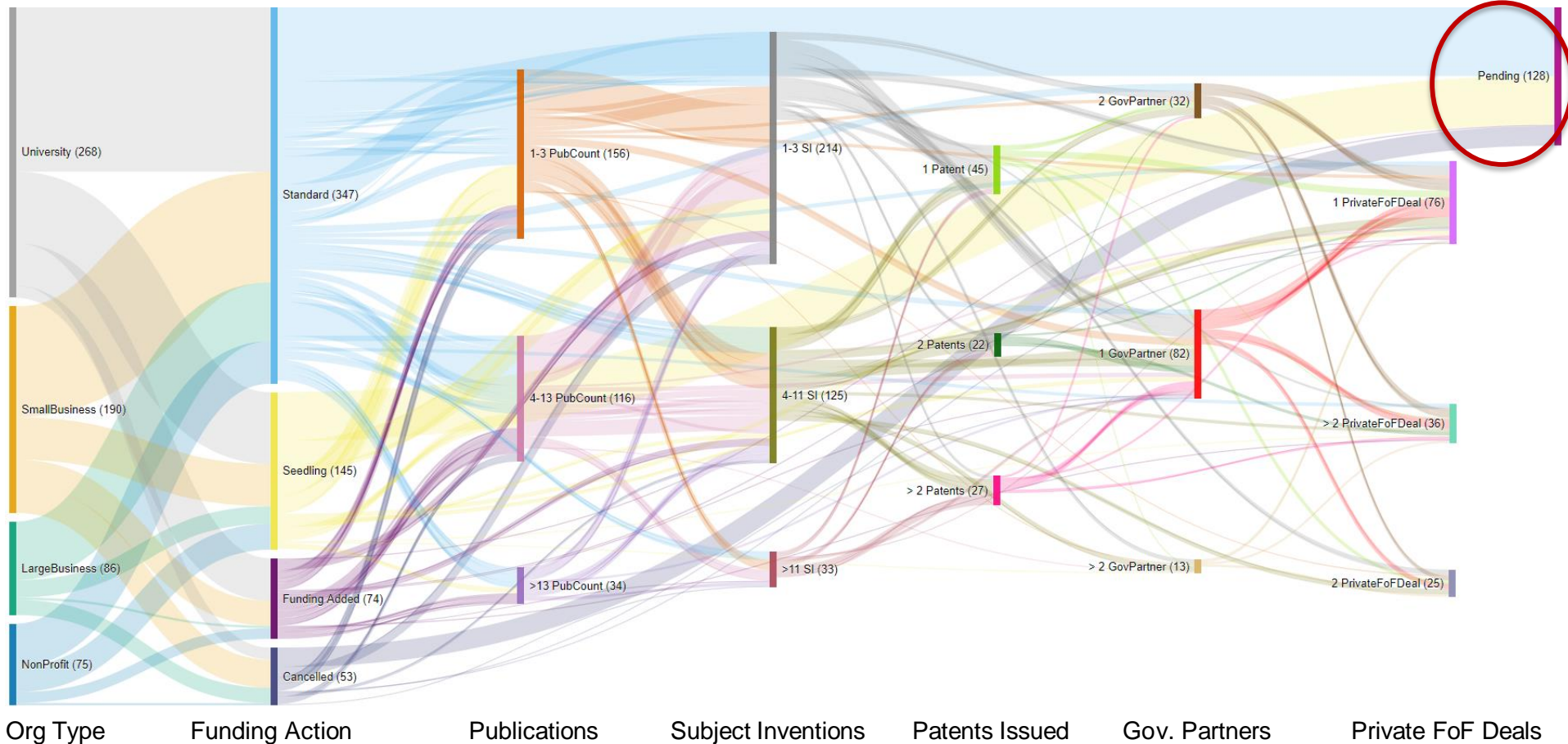
**Step 2:** Choose the type of analysis you would like to conduct from the “Analysis by” drop-down and then the “Selection” drop-down. Upon making a selection, all interactive assets of the dashboard will be updated accordingly. By default, all dashboard assets are rendered based on the data from the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

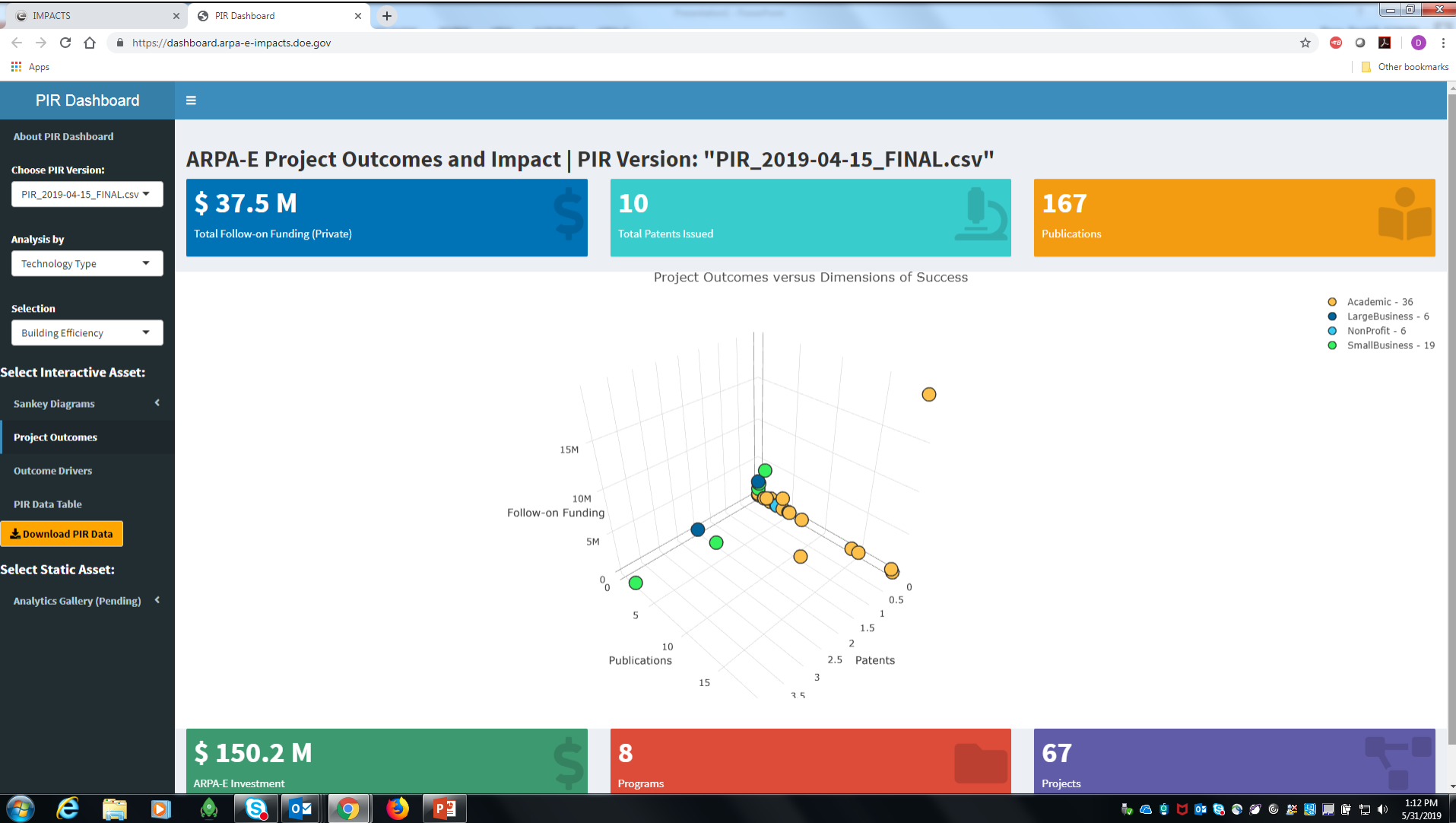
**Step 4 (Optional):** Click the “Download PIR Data” button to download either the full PIR dataset or a subset of the PIR based on current analysis. A .xlsx file will be downloaded to your download folder, with the filename beginning with ‘PIR\_AnalysisExport\_’. This file will also contain a data dictionary 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.

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



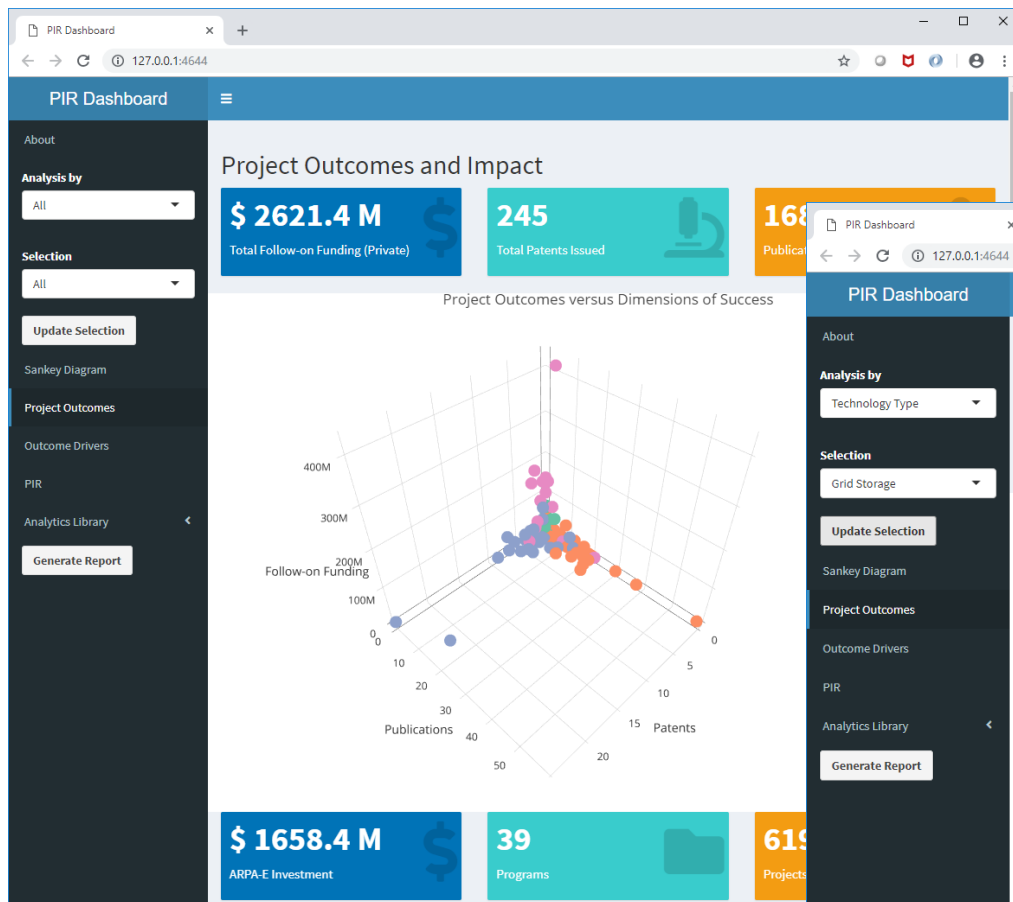
# Advanced Data Reporting: Clusters



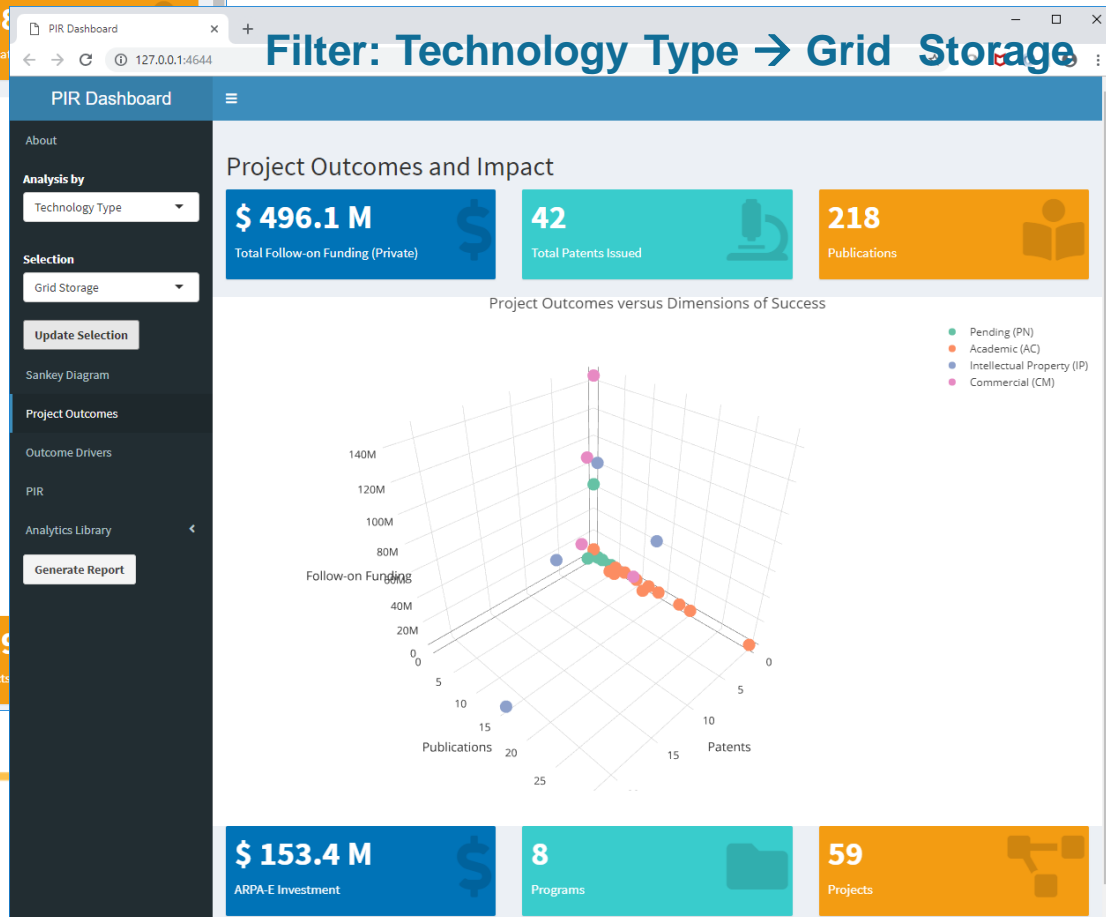


# Advanced Data Reporting: Filters

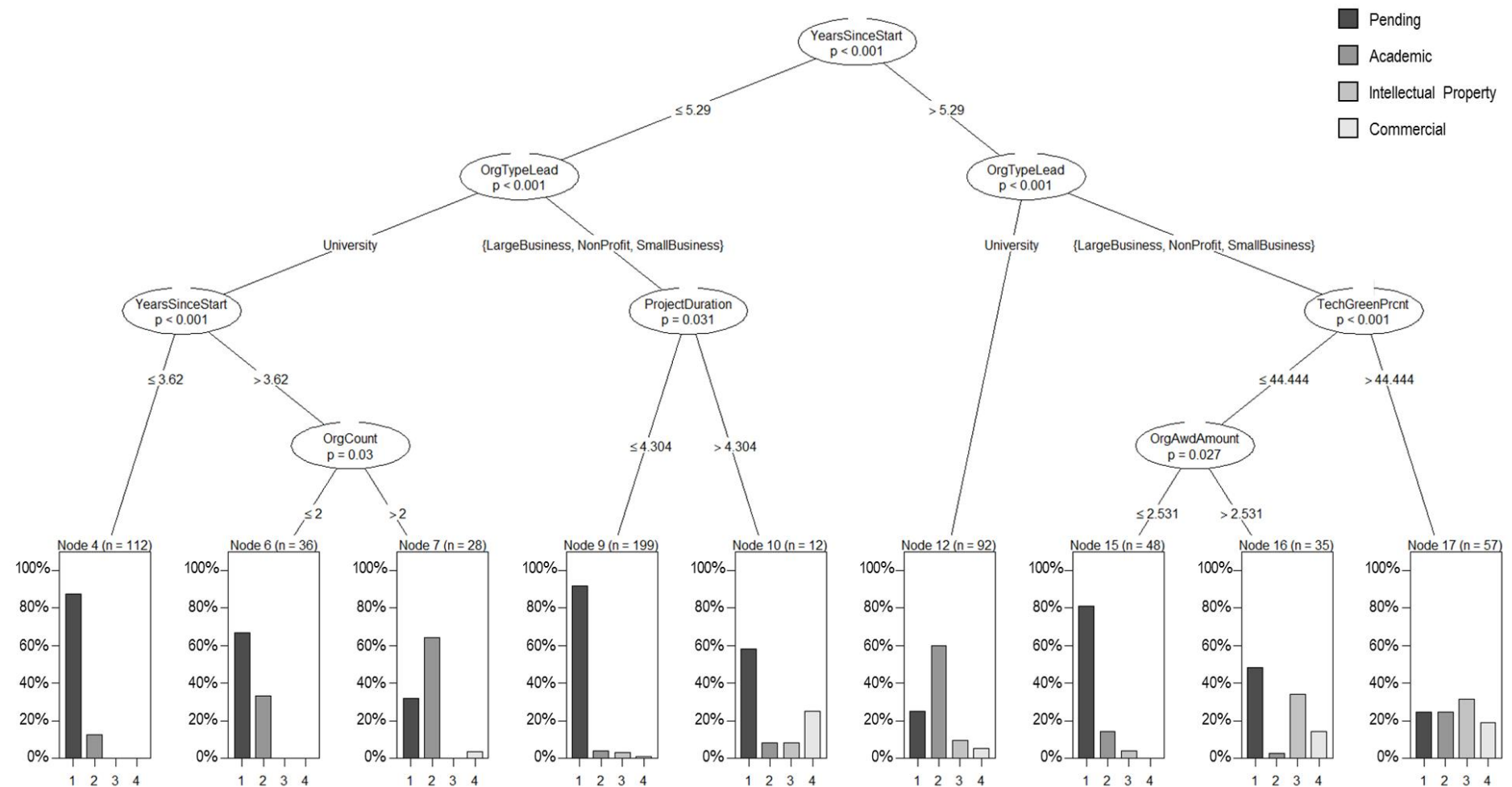
Enabling analyses by Program, Technology Type and Project Status



Filter: Technology Type → Grid Storage



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





# Full Text Search: Landing Page

The screenshot shows a web browser window with multiple tabs. The active tab is titled "IMPACTS" and the address bar shows the URL "https://apps.arpa-e-impacts.doe.gov/solr/". The browser's address bar also displays "Apps" and "Other bookmarks". The page content features the ARPA-E logo in the top left corner and navigation links for "Home", "Getting Started", and "About" in the top right corner. The main area of the page is white and contains the Solr logo, which consists of the word "Solr" in a black serif font and a red circular icon with white radial lines to its right. Below the logo is a search input field with the placeholder text "Search with Solr" and a blue "Search" button. The footer of the page is dark gray and contains the following text: "Advanced Research Projects Agency - Energy", "U.S. Department of Energy", "1000 Independence Ave SW - Washington DC 20585", "Questions or comments?", and "Contact: ARPA-E-IMPACTS-Support@hq.doe.gov". A small "NOTICE TO USERS" is also present in the footer, stating that the system is a Federal computer system and is the property of the United States Government. The Windows taskbar is visible at the bottom of the screen, showing various application icons and the system clock indicating 1:24 PM on 5/31/2019.

arpa·e

Home Getting Started About

Solr

Search with Solr

Search

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 or unauthorized) have no explicit or implicit expectation of privacy. Unauthorized or improper use of this system may result in administrative disciplinary action and civil and criminal penalties. By continuing to use this system you indicate your awareness of and consent to these terms and conditions of use. CLOSE THIS WINDOW  
IMMEDIATELY if you do not agree to the conditions stated in this warning.

# Full Text Search: Results

IMPACTS x IMPACTS x Lingo4G Explorer x IMPACTS x PIR Dashboard x +

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

Apps Other bookmarks

arpa-e Home Getting Started About

Solr energy Search

**Effective Selective Area Doping for GaN Vertical Power Transistors Enabled by Innovative Materials Engineering**  
Program: PNDIODES | FOA Number: DE-FOA-0001691 | Control Number: 1691-1502  
Abstract: No abstract...  
POC: Yuji Zhao | Document Type: Quarterly Report | Status: Active

**Demonstration of PN-junctions by Ion implantation techniques for GaN (DOPING-GaN)**  
Program: PNDIODES | FOA Number: DE-FOA-0001691 | Control Number: 1691-1517  
Abstract: No abstract...  
POC: Fatemeh Shahedipour-Sandvik | Document Type: Quarterly Report | Status: Active

**Effective Selective Area Doping for GaN Vertical Power Transistors Enabled by Innovative Materials Engineering**  
Program: PNDIODES | FOA Number: DE-FOA-0001691 | Control Number: 1691-1502  
Abstract: No abstract...  
POC: Yuji Zhao | Document Type: Quarterly Report | Status: Active

**GaN power devices enabled by innovations in implantation and regrowth**  
Program: PNDIODES | FOA Number: DE-FOA-0001691 | Control Number: 1691-1511  
Abstract: The Cornell-NREL-NREUnipress team proposes to achieve mechanistic understanding of defect

<< prev 1 2 next >>

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 or unauthorized) have no explicit or implicit expectation of privacy. Unauthorized or improper use of this system may result in administrative disciplinary action and civil and criminal penalties. By continuing to use this system you indicate your awareness of and consent to these terms and conditions of use. CLOSE THIS WINDOW  
IMMEDIATELY if you do not agree to the conditions stated in this warning.

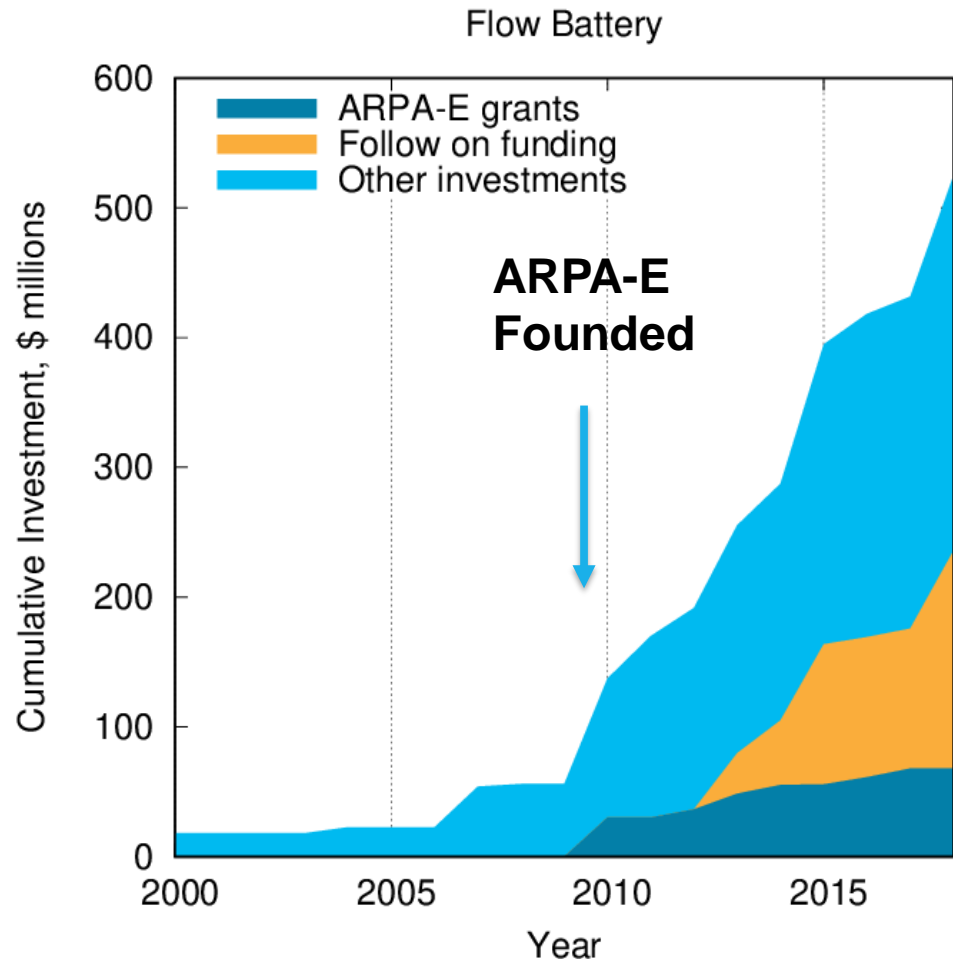
1:25 PM 5/31/2019



Before, During and After Program Launch

# CASE STUDIES

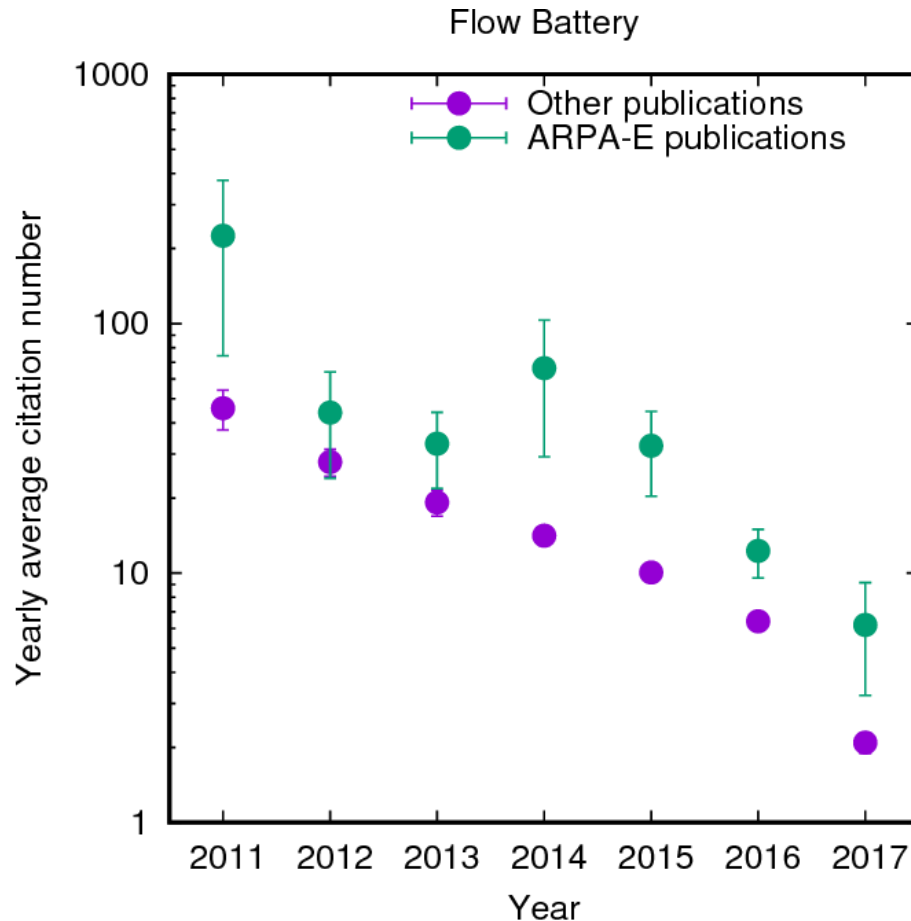
# ARPA-E grants spurred industry funding



“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

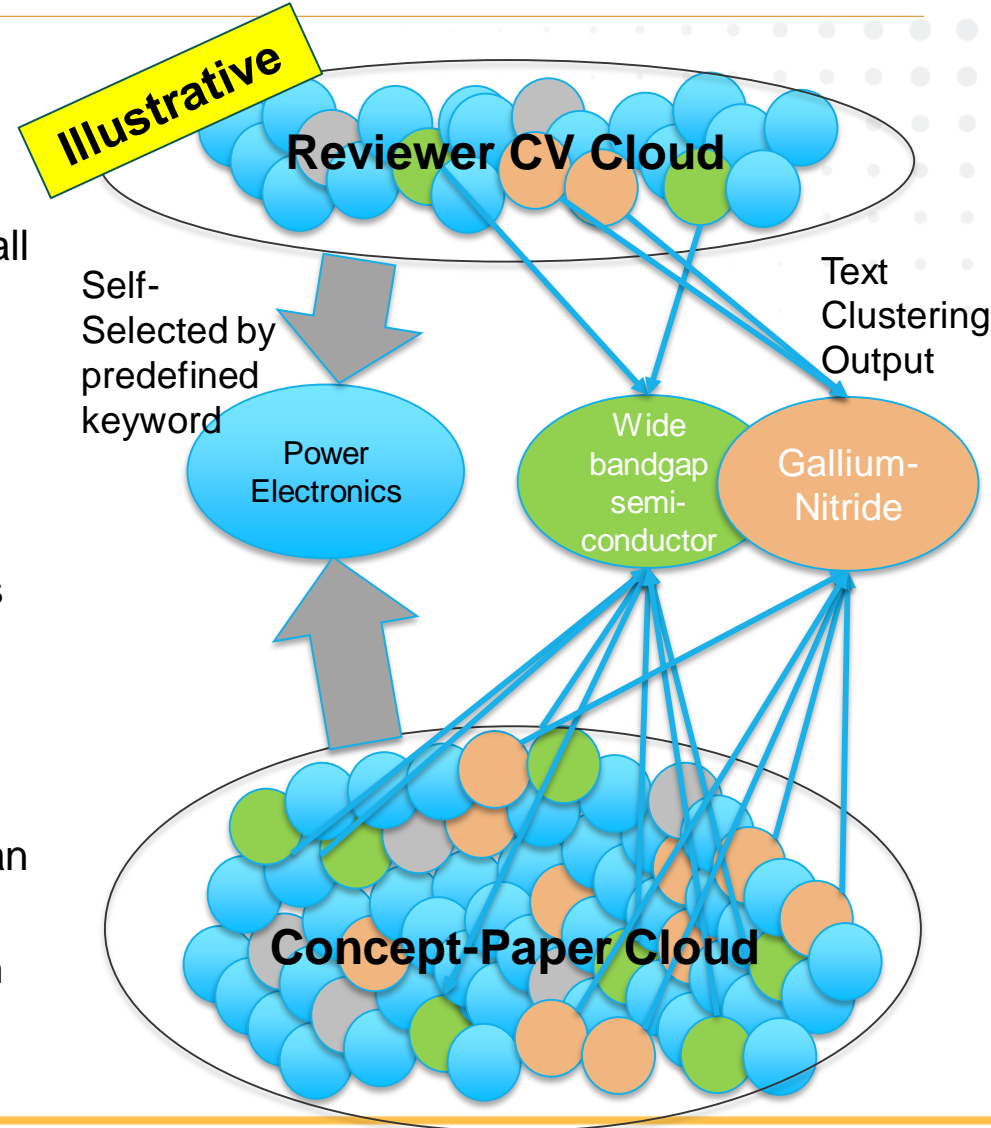
# Case Study: Leveraging IMPACTS for Concept Paper Reviewer Match

## ► 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?



U.S. DEPARTMENT OF  
**ENERGY**

<https://arpa-e.energy.gov>