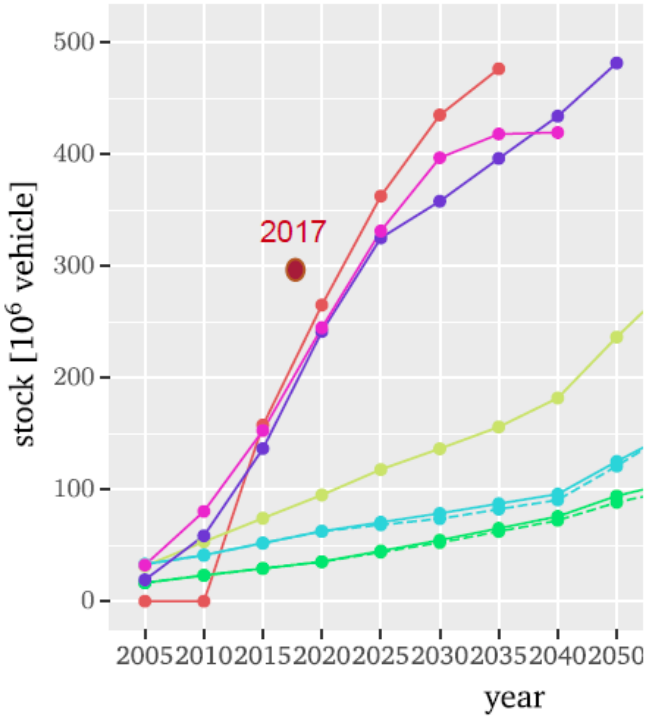
KAPSARC
مركز الملك عبدالله للدراسات والبحوث البترولية
King Abdullah Petroleum Studies and Research Center

Value of data transparency

from an energy researcher perspective

Amar Amarnath, Chief of Data Management
Dec 2018

Researcher challenge example #1 : Trust on research output



- Huge uncertainty about China: China's LDV stock
- Will there be 90 million cars or 500 million cars in China by 2050?
- Large implications on vehicle stock, fuel prices, urban planning, refineries investment, etc.

Researcher challenge example #2: Variation in our feedstack - data

How Much Do People Travel? How Certain Are We?

Thousand KMs/capita/year, all modes

	Australia	Brazil	China	U.S.
Energy Modeler 1	26.2	5.4	5.0	23.1
Energy Modeler 2	21.7	8.3	8.4	26.6
Energy Modeler 3	33.7	6.9		15.2
Energy Modeler 4	43.5	5.4	4.5	27.7
Energy Modeler 5		4.2	6.3	19.7
Energy Modeler 6	17.5	8.4	6.5	26.8

Researcher challenge example #3: Takes too long, missing monthly

Challenges Faced and Action Taken (I)

It took at least 3 months to finalise the energy consumption and CO_2 data that involved numerous email/telephone exchanges with the admin staff and the people employed working on the data due to the following:

- Historical end points were incorrect for a number of series
- Some of the series were interpolated from annual figures and the utilised procedures had produced negative values
- Missing values introduced in error
- No documentation of data sources, definitions and procedures (including interpolation method and factors used to constuct CO_2 emissions)
  - In the end we had to drop CO_2 emissions because it was simply constructed as a weighted average of the consumption data provided and hence the two were highly correlated

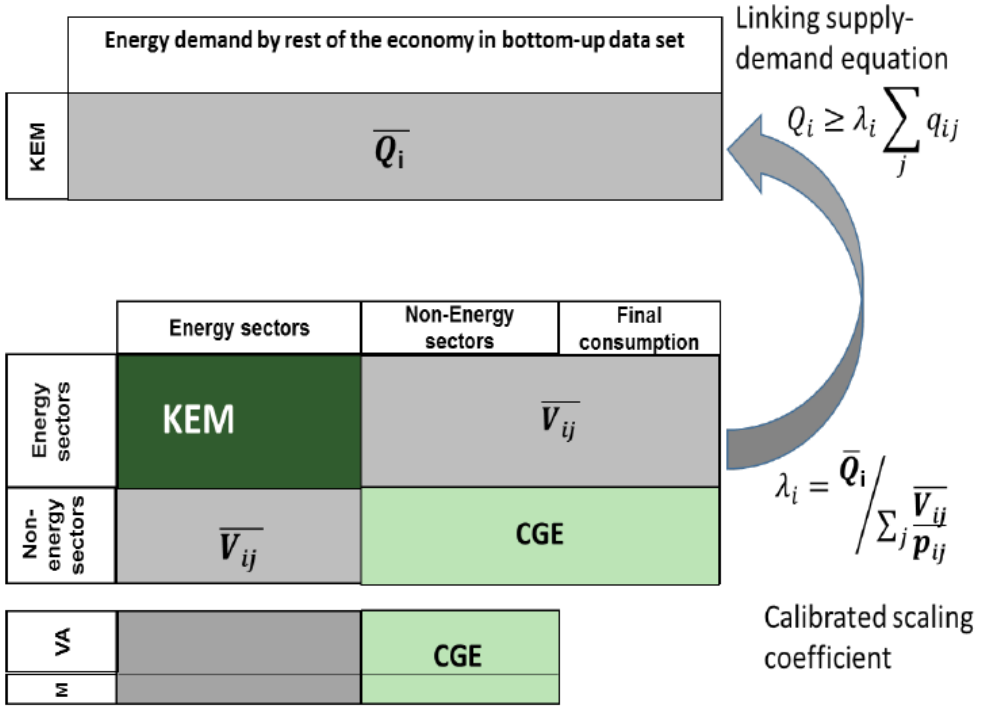
Challenges Faced and Action Taken (II)

- For energy prices quarterly coverage was very poor (Datastream) or included many gaps and a lot of missing series (International Energy Agency)
  - Had to use world energy prices instead of country-specific prices
- The energy consumption data was not temperature adjusted
  - Spent 3 months to download and construct degree days from temperature data at the quarterly frequency
  - KAPSARC's global degree days database for energy-related applications covering 147 countries over a period of several decades could have been useful for our purposes, which I recently became aware of

Researcher challenge example #4: Sector level data, granularity

KEM – G: Data challenges

- **Data availability:**
  - Social Accounting matrix
  - Disaggregated data (energy consumption, sectors)
- **Inconsistencies between data sets**
  - But top-down and bottom-up data don't always match…
- **Data management**
  - Update the calibration year
  - Linking the model to other energy models (KEM, KEM-GCC,…)



Dealing with the data differences: Example of domestic energy demand

Researcher challenge example #5: monthly data please!

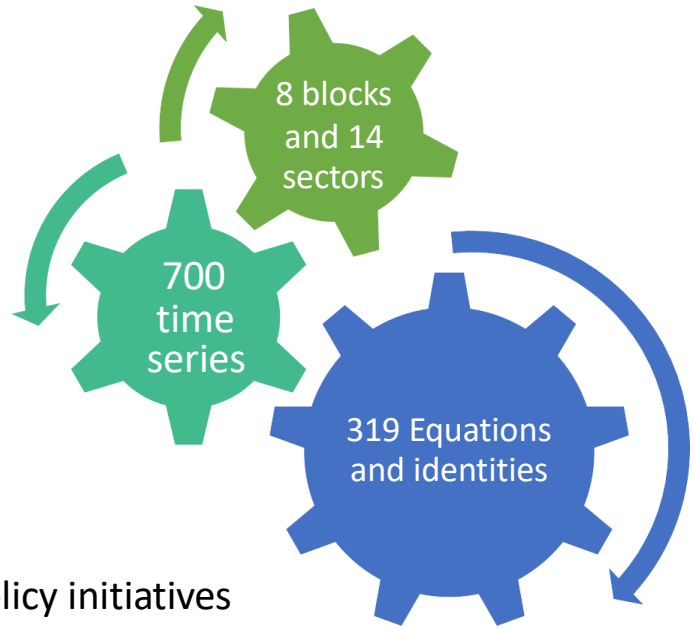
- Large-scale administrative data sets and proprietary private sector data can greatly improve the way we measure, track, and describe economic activity.
- Models are constructed under severe constraints of limited availability of data, mostly compiled annually

Serious reduction in degrees of freedom and diagnostic testing.

- Limited policy relevance of model results due to lack of disaggregated monthly, quarterly data
- Lack of disaggregated regional data
- The choice for appropriate lag length is a challenging due to the shortness of annual time series and the relative absence of monthly and quarterly data for most macro-variables.

KAPSARC Global Energy Macro-econometric Model

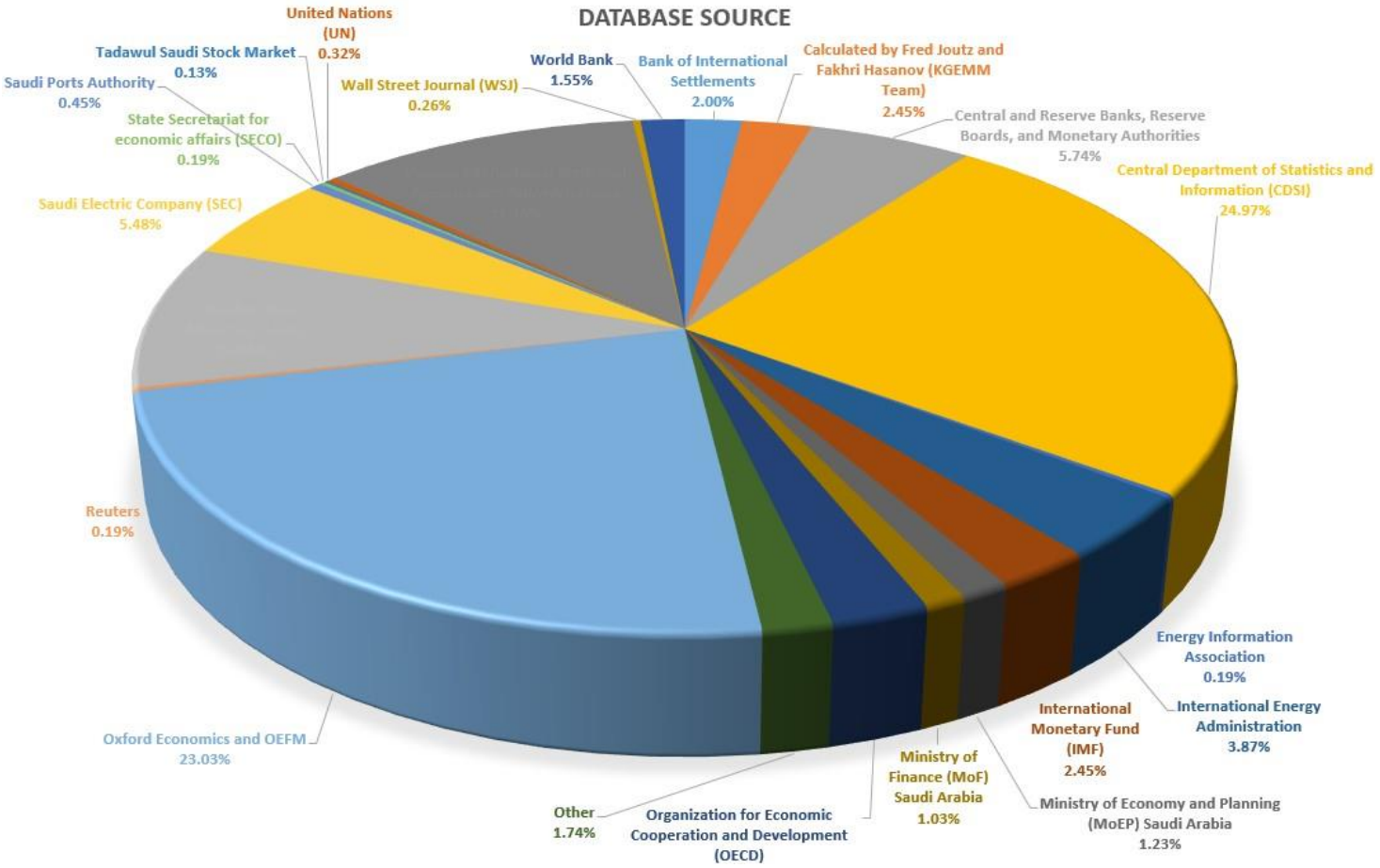
- **Objective**
  - A policy analysis tool examining the impacts of policy decisions and the interaction between the global economy and macroeconomic energy environment of Saudi Arabia.



- **Use**
  - To evaluate the impacts of different policy initiatives and macroeconomic targets
  - Analyzing and forecasting behaviors of energy and macroeconomic indicators

Models' data sources

*There are more than **700** time sries aggregated and disaggregated energy and economic variables from **22** sources as of 2018*



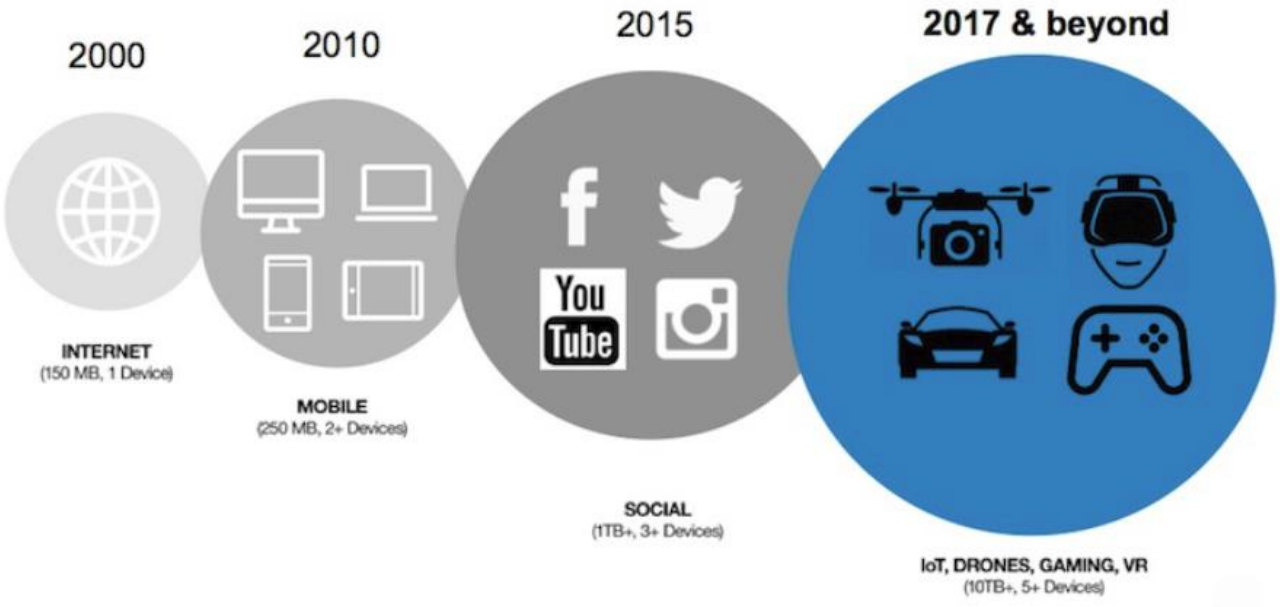
Researcher Data Challenges

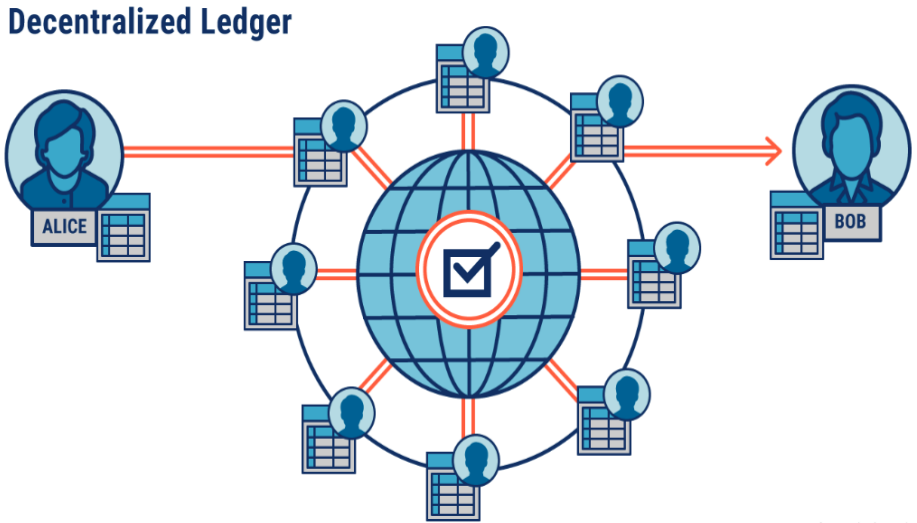
- Data revision without alerts
- Archive old data not publically available
- Unavailability of disaggregated data
  - Sectoral investment
  - Sectoral employment by nation and gender
  - Energy consumption and prices by sector and customer segments

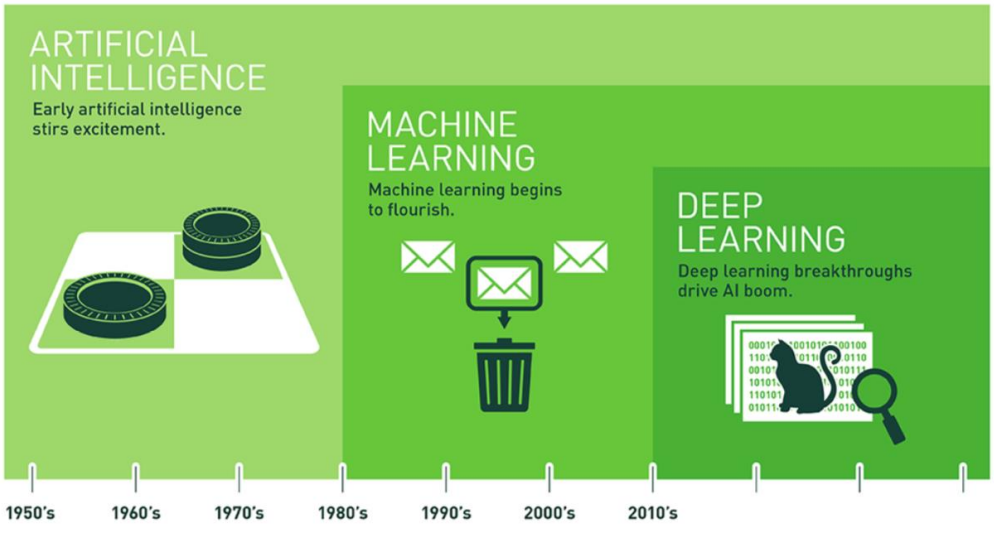
Consequences without official statistical data

- Out of date representation of the country economy
- Absence/improper of granular relationships in the economy
- Unable to conduct short-term policy analysis and projections
- Unable to represent the economy in a meaningful way

A minute into future where data and prediction landscape…







The King Abdullah Petroleum Studies and Research Center (KAPSARC) is a non-profit global institution dedicated to independent research into energy economics, policy, technology and the environment, across all types of energy.

KAPSARC's mandate is to advance the understanding of energy challenges and opportunities facing the world today and tomorrow, through unbiased, independent, and high-caliber research for the benefit of society. KAPSARC is located in Riyadh, Saudi Arabia.

Board of Trustees

H.E. Eng. Khalid Al Falih
Chairman of the Board

H.E. Prof. Hashim Yamani
President of the King Abdullah City for Atomic and Renewable Energy (KA-CARE)

Prof. Sir Keith O'Nions
Chairman of Cambridge Enterprise
Former Rector of Imperial College

Prof. Robert Brown
President of Boston University

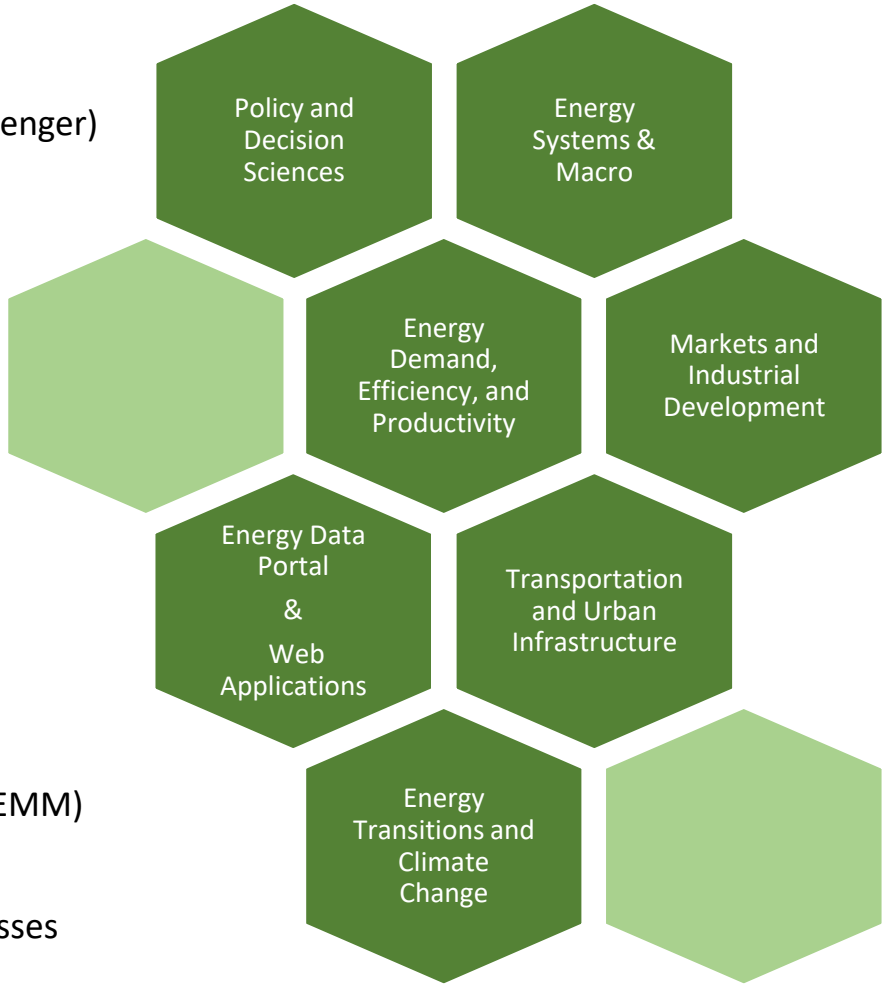
H.E. Dr. Muhammad Al-Jasser
Advisor at the Royal Court,
Former Minister of Economy and Planning

Adam Sieminski
President of KAPSARC

Ultimate authority of KAPSARC
Sets policies and approves strategic, and operational plans of the Center

Research Program Areas

- Future of transportation fuel demand (freight and passenger)
- Future of global oil and natural gas markets
- Electricity sector transitions
- Climate change policies and Energy technologies
- Policy and decision science

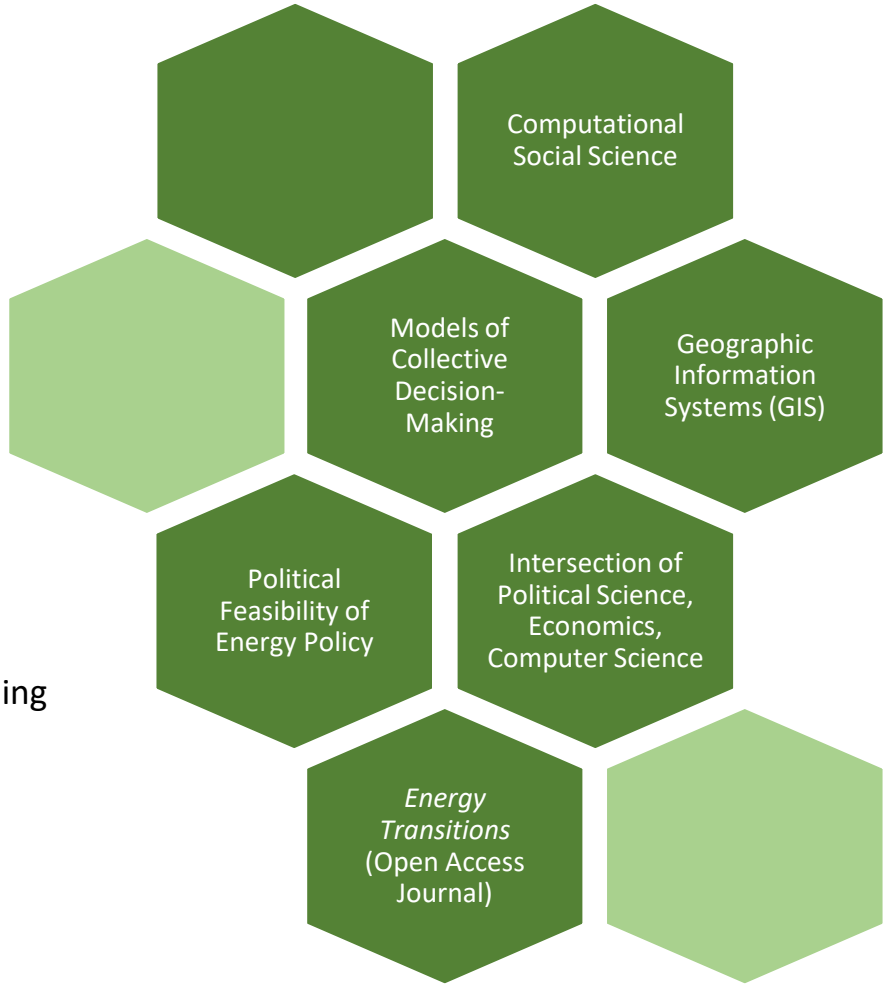


Research Platforms

- OpenKAPSARC
- Energy data portal, data insights and web apps
- KAPSARC Global Energy Macroeconometric Model (KGEMM)
- KAPSARC Energy Model (KEM)
- KTAB toolkit – models collective decision making processes

Policy and Decision Science Program

- What does economics not explain about the world?
- Computational social science as a way to understand complex social-political phenomena
- Explicit focus on how groups of people make decisions and influence each other
- Geospatially study people, communities, and cultures linked to place and space (human geography)
- Ensure real world relevance to decision-makers

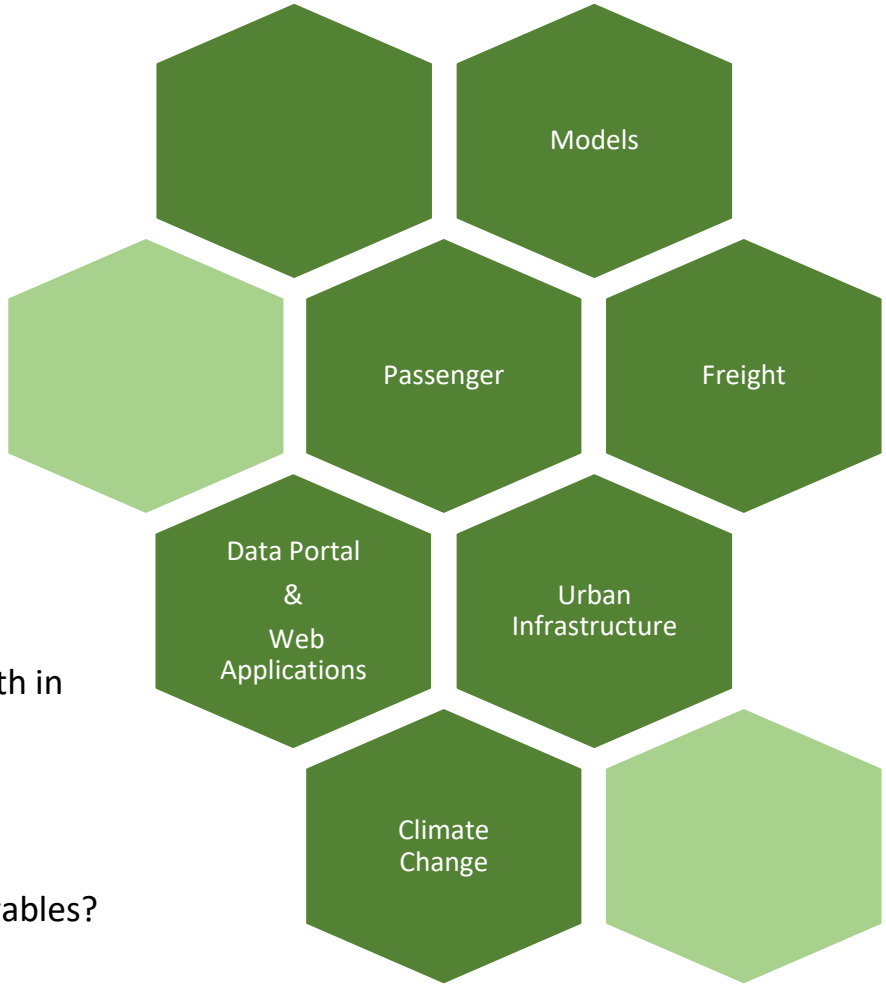


A Sampling of our Research

- KAPSARC Toolkit for Behavioral Analysis (KTAB)
- Explain household decision-making with machine learning and KTAB modules
- Statistical evaluation of political and social factors that impact energy markets
- Political will of GCC countries to enact policies for an energy transition
- Political feasibility of implementing and extending the Paris Agreement

Energy Efficiency and Productivity

- The economics of energy efficiency
- Energy productivity
- The impact of energy prices
- The rebound effect in transportation
- Energy efficiency in building
- Energy and its impact on employment



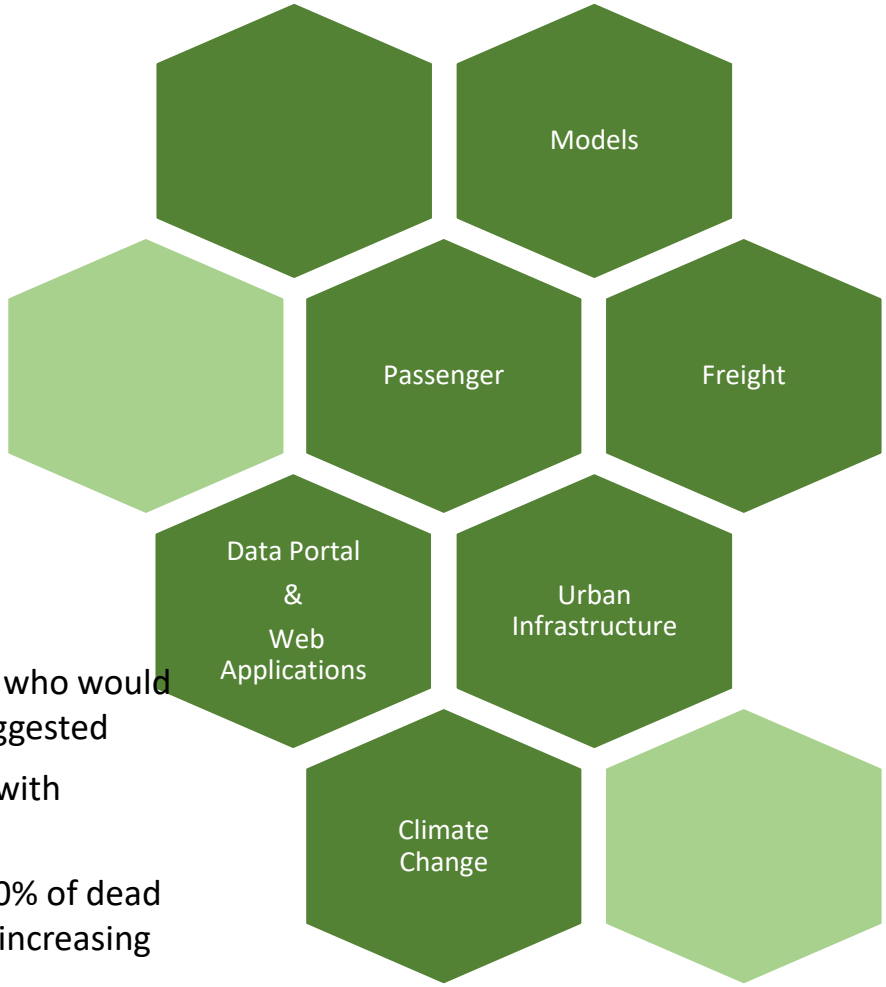
Research Questions

- How can countries in the GCC region manage the growth in energy demand?
- What effect could energy price reform have?
- What are the multiple benefits of energy efficiency?
- What is the potential for employment in energy renewables?

Transportation

Goal: Meeting the increasing mobility demand while minimizing energy consumption

- ❑ Contextual view
- ❑ Technology imperatives
- ❑ Economic viability
- ❑ Policy support

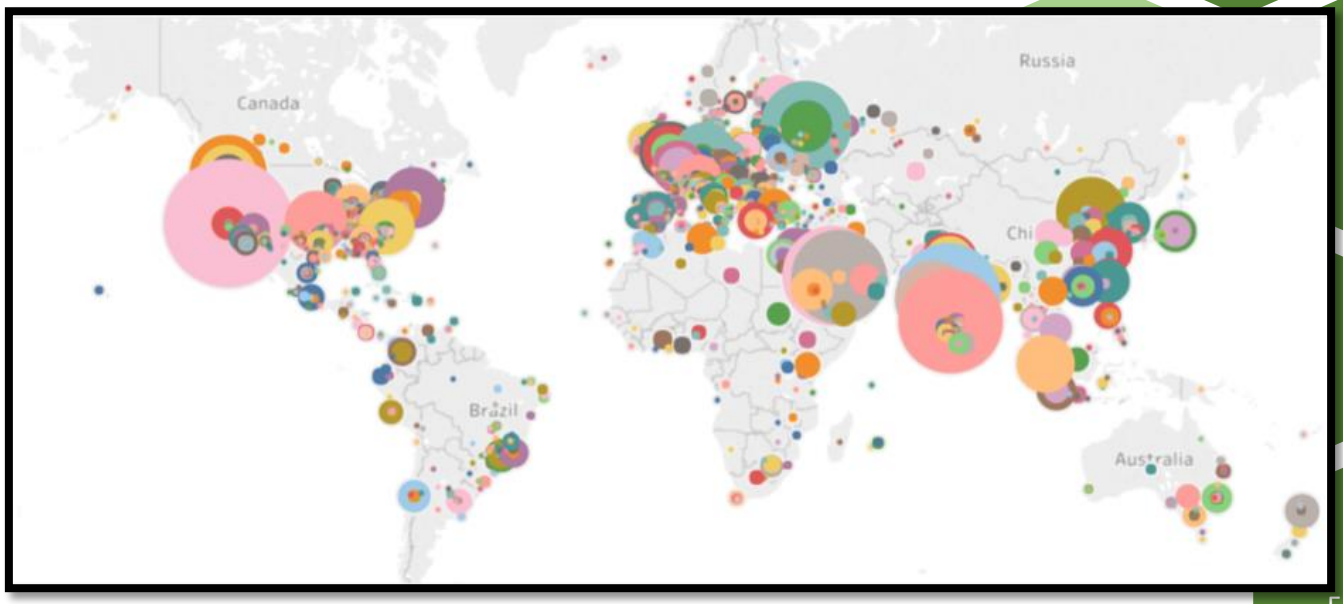


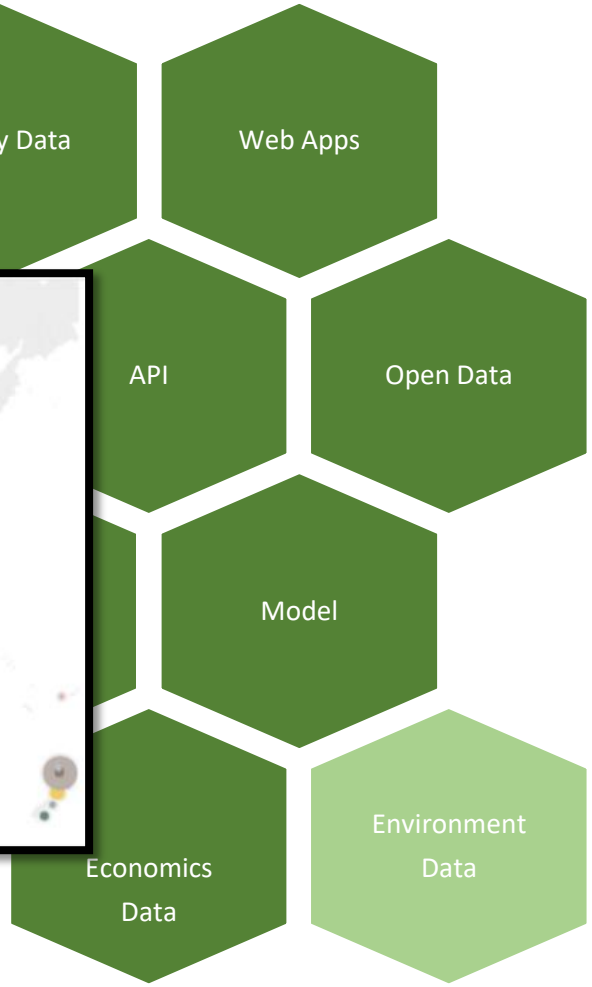
Eg. Insights from study

- ❑ 80% of subsidies on EV purchase are utilized by people who would have bought EVs anyway (US data) – optimal design suggested
- ❑ Freight demand shifting spatially and commodity wise with increasing economic growth in China
- ❑ Uberization of freight can substantially decrease the 40% of dead mileage in truck movement in China and India thereby increasing energy efficiency

Energy Information Management - Data, Web Apps, Models

- Develop web portal hosting “data, models and algorithms”
- Acquire, wrangle and delineate data from models
- Aggregate and make machine readable open data to world
- Publish data insights showcasing changes and trends
- Develop web applications to showcase models

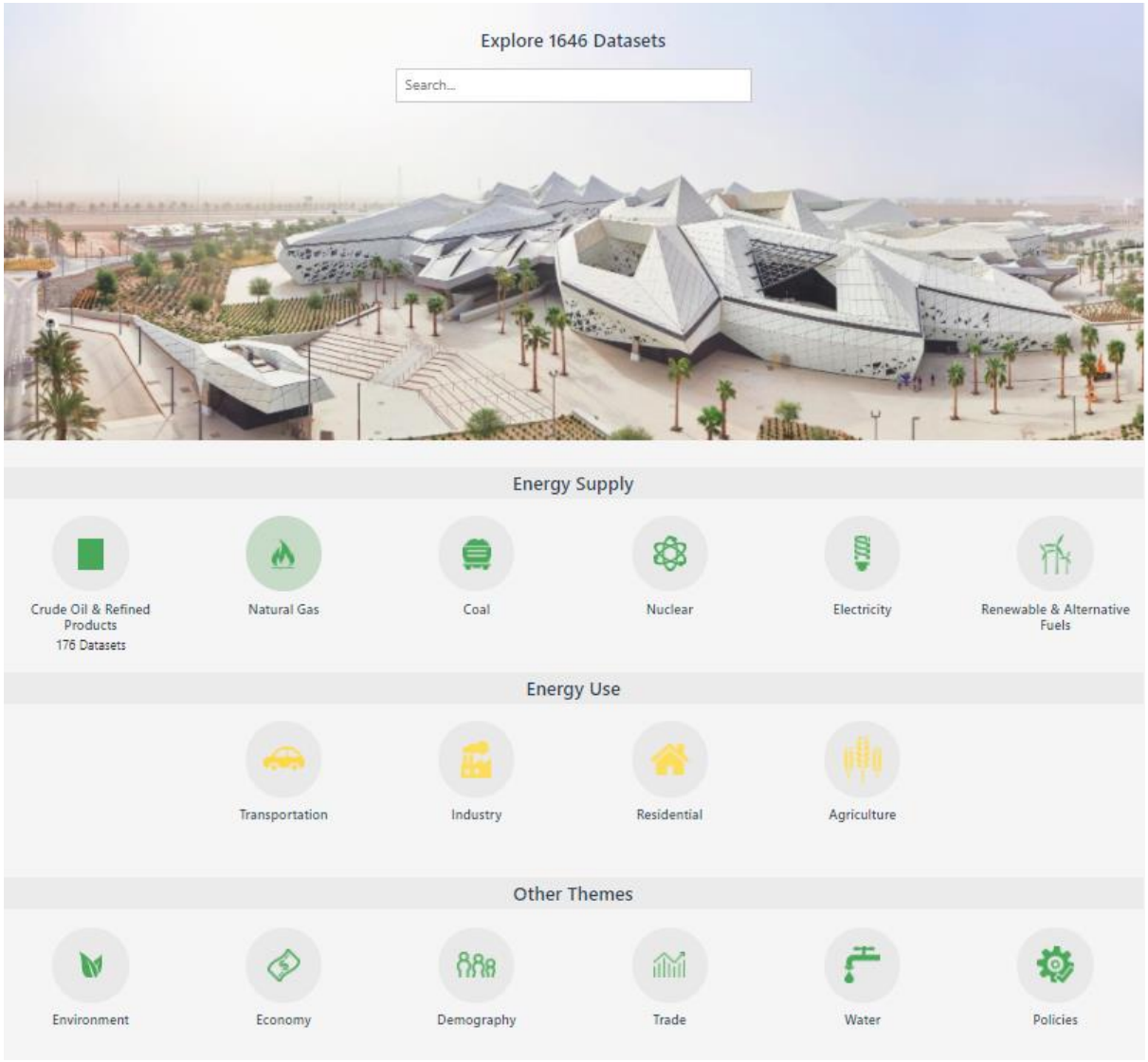




- 30M records, 3K indicators, 10 years time series data
- 45 days average data currency from 150 data sources
- 6000 searches and 40 downloads per day

OPEN KAPSARC www.kapsarc.org https://datasource.kapsarc.org

features data and web applications to run scenarios on KAPSARC models.



KAPSARC OpenTools

English العربية Subscribe

KAPSARC

Home Researchers Published Articles OpenKAPSARC News & Events About Careers

OpenKAPSARC

The OpenKAPSARC platform provides free access to KAPSARC's data and research tools, to advance the understanding of energy economics and environment policy. It contains over 1000 machine-readable (API ready), model-ready datasets, 3000 indicators and over 30 million records.

Web Apps | Developer Tools | Data Insight

KAPSARC Energy Data Portal

Go To App | Details

KAPSARC Insights

Go To App | Details

KAPSARC Maritime Transport Analysis Framework

Go To App

KAPSARC Building Energy Assessment Tool

Go To App | Details

KAPSARC Global Energy Macroeconometric Model

Go To App | Details

KAPSARC Energy Model for Saudi Arabia

Go To App | Details

Researcher data challenges aligned to model and insight development

Data quality

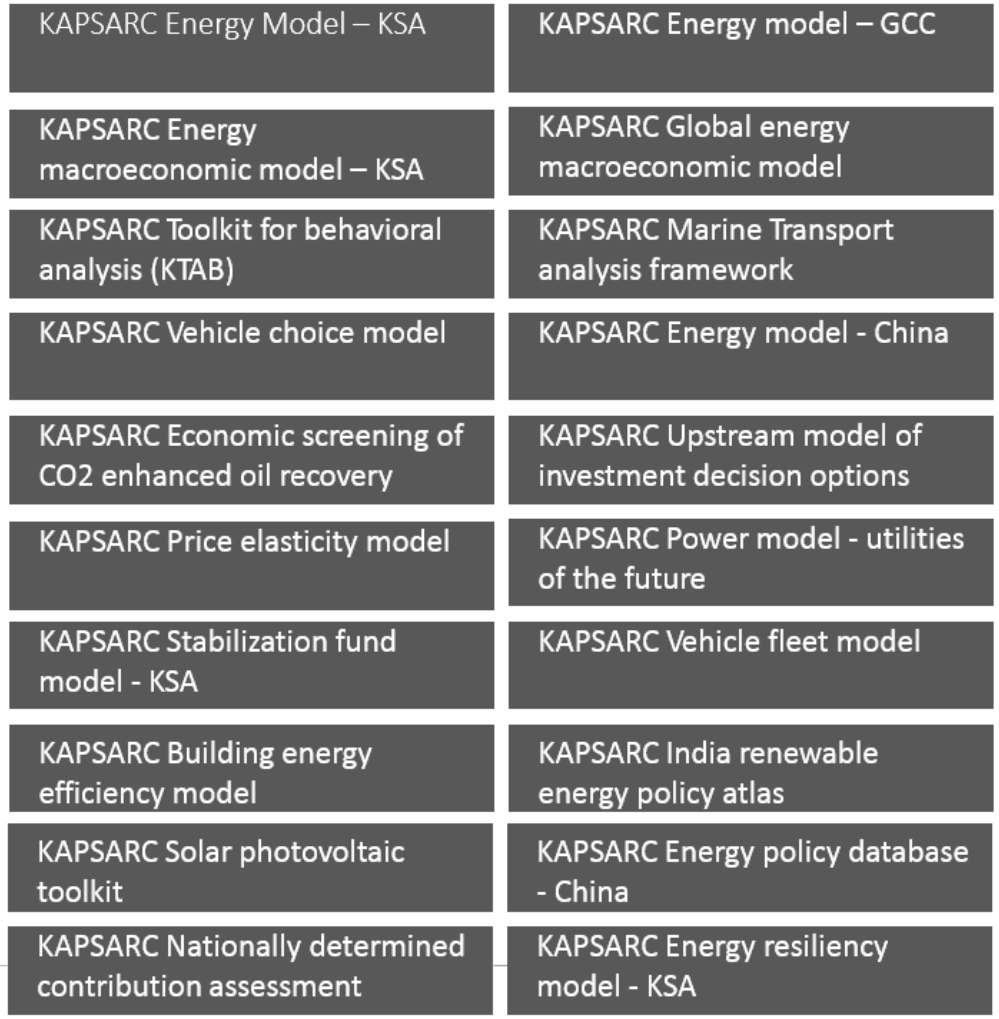
- Currency (monthly PLEASE!)
- Completeness
  - Coverage
  - Temporal
  - Granularity (disaggregated regional data)
- Consistency
- Correctness

Data trust improves when data is

- Traceable back to OFFICIAL source
- Used, reviewed, commented and critiqued
- Relevant to researcher context
- Interpretable across systems
- Definitions are standard, meta data published

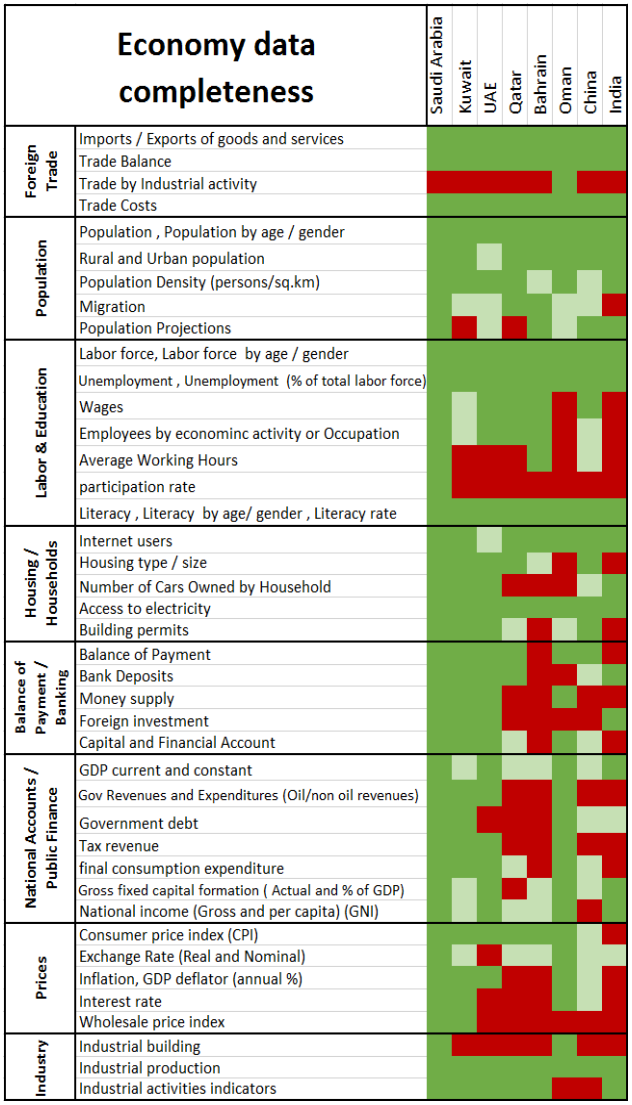
Security

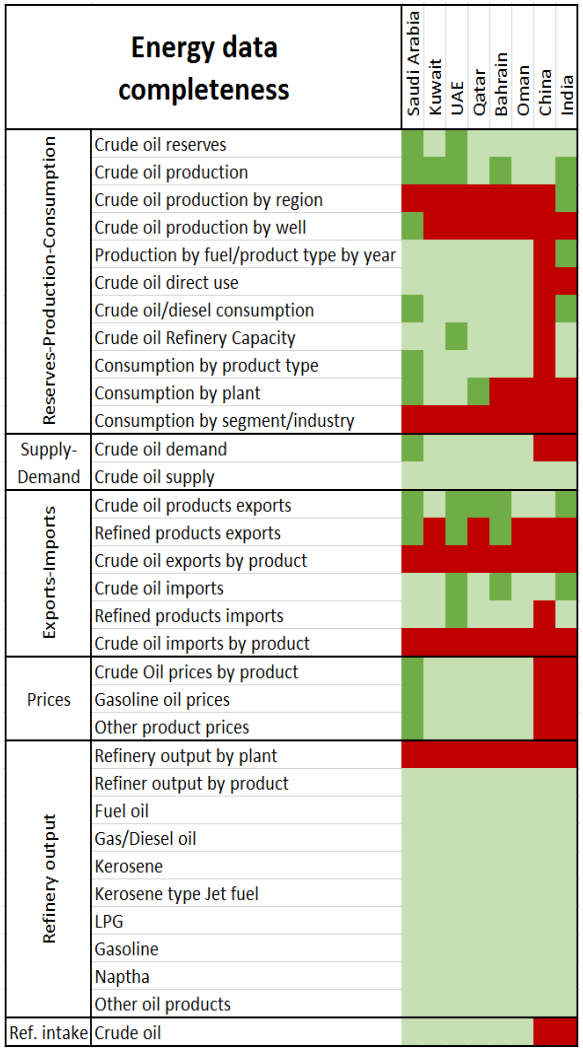
- Public, Restricted. Confidential

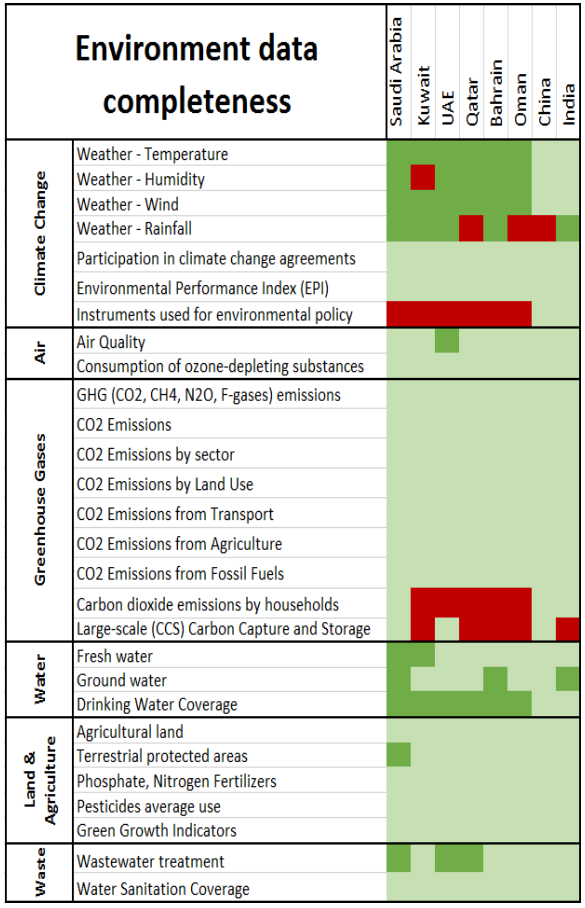


Granularity is key, better data, better insights

Regional Source	International Source	Not available



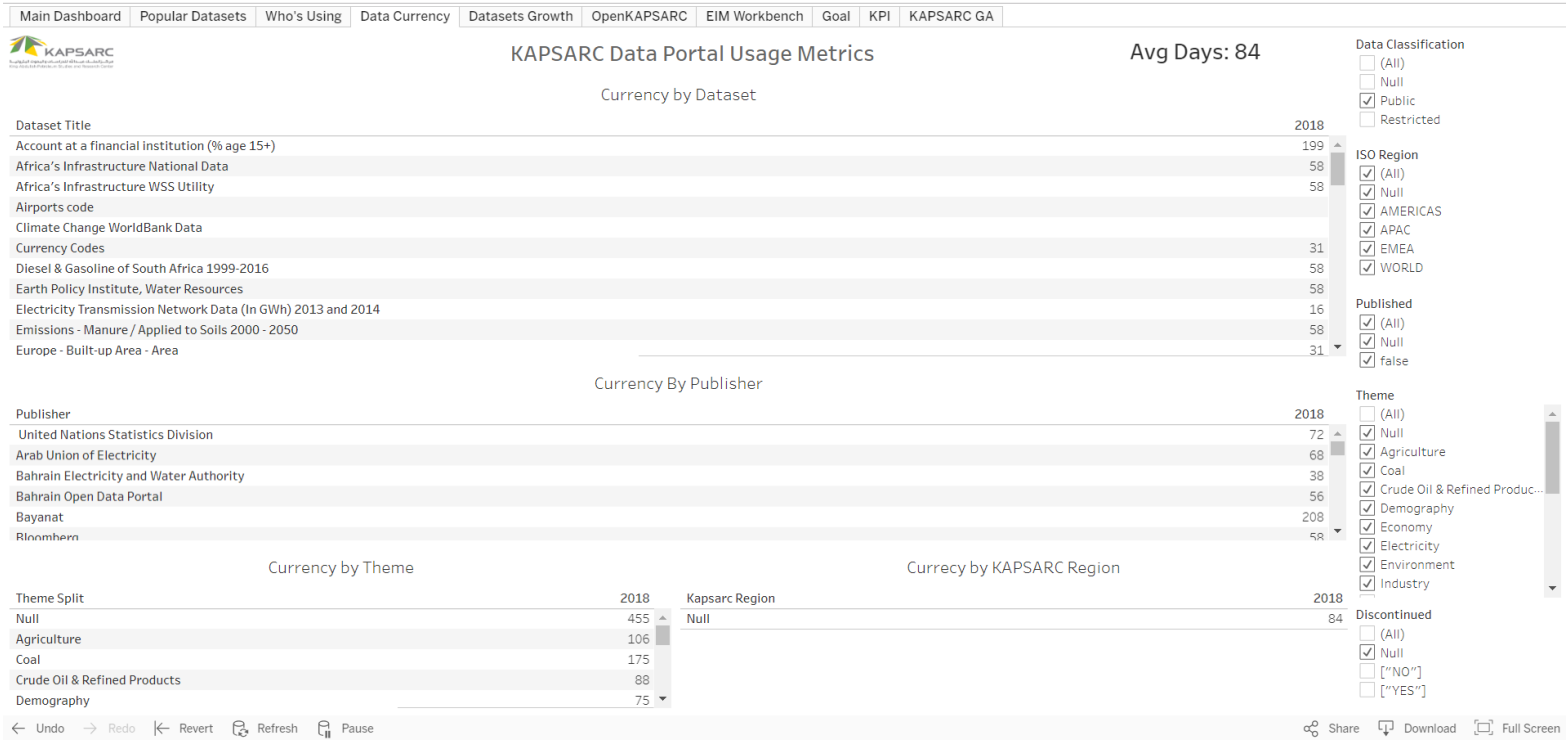




92% Annual, 6% Quarterly
70% Data 2015 – 2018

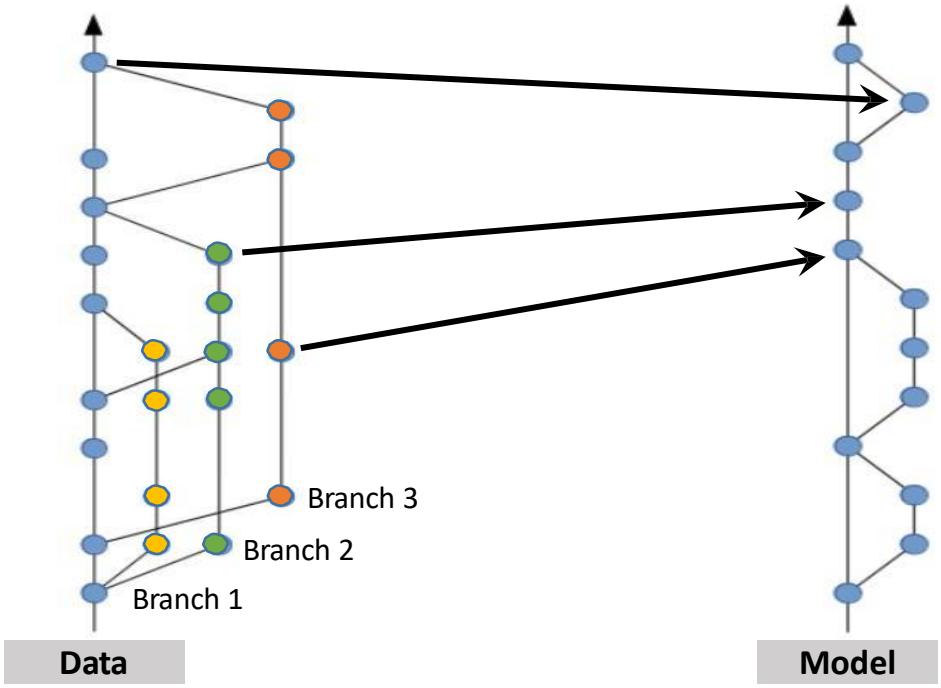
Data Management Key Performance Indicators

KPI Measures	2018 Targets	Year to Date	Oct 2018	Nov 2018	Trend
1. KAPSARC data portal external users' actions	2M	1.9M	228K	142K	Down
2. Data portal external users' downloads	15K	12.4K	1800	968	Down
3. Number of insight web apps developed	6	7	1	1	Up
4. # of model runs on OpenKAPSARC	300	296	36	30	Down
5. # of data insights	10	18	2	15	Up
6. # of models with open data	6	8	3	3	Neutral
7. Portal Data currency	45	45	45	84	Down



KAPSARC Model Data Editor

Addressing challenge of data versions



1.Seed Data Automated

2. Edit Data

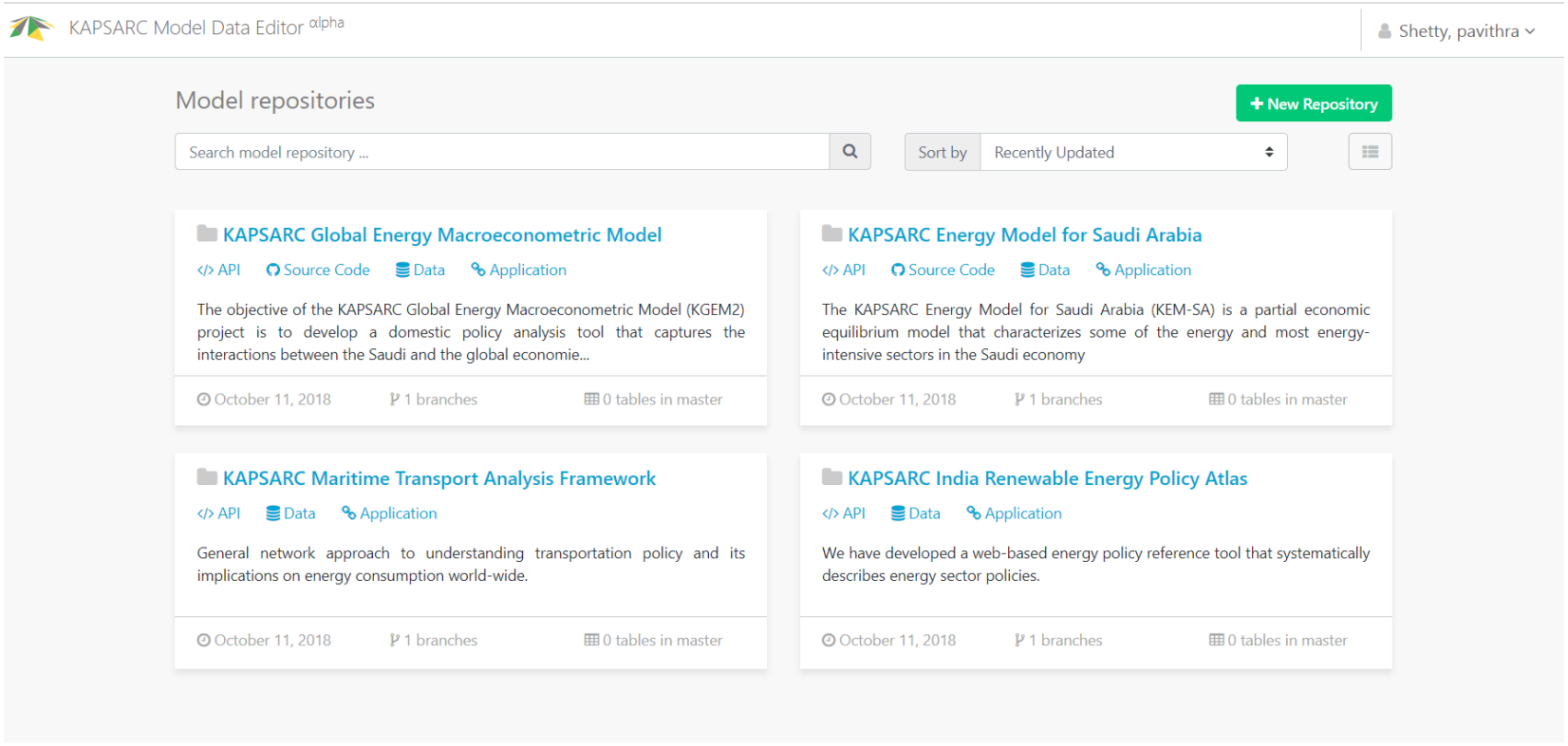
3. Version Alert

4. Collaborate

5. API Calls Hist.

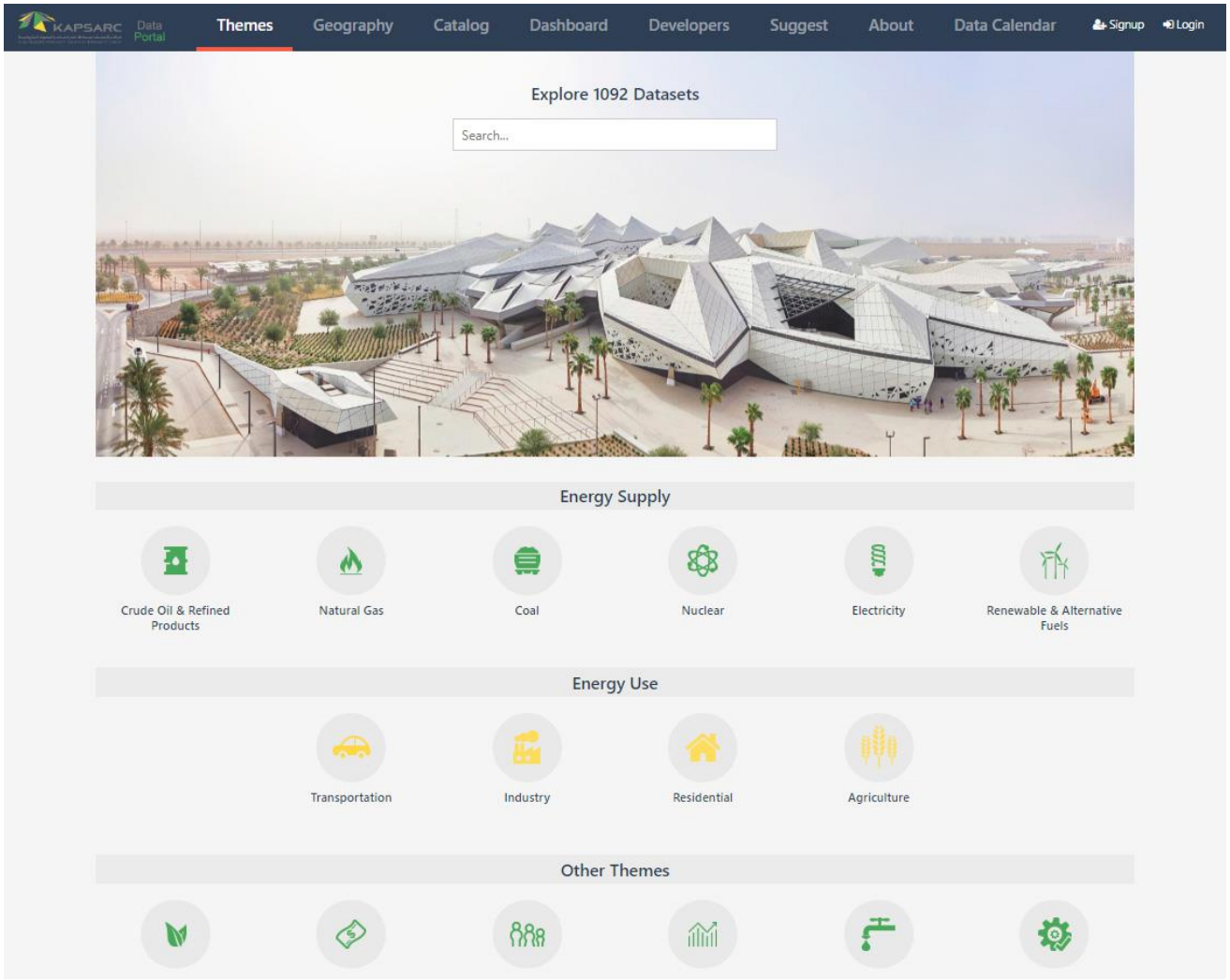
KAPSARC DataHub for Modelers

Open source portal to manage models' data and call via APIs (Application Programming Interface)



KAPSARC OpenData

In 2018 datasource.kapsarc.org served over 2,000,000 searches and 15000 downloads in from 100 countries



1K Datasets
machine readable

30M Records

45 Days Currency

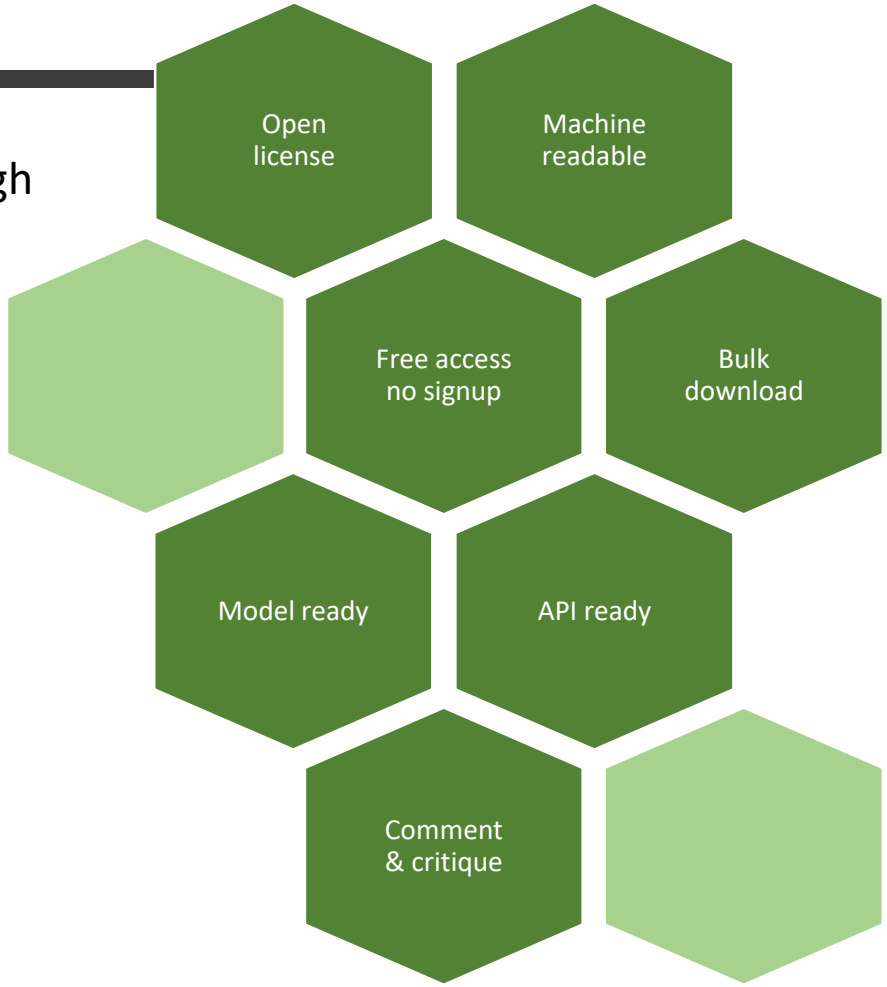
150 Data Sources

60 GCC **Sources**

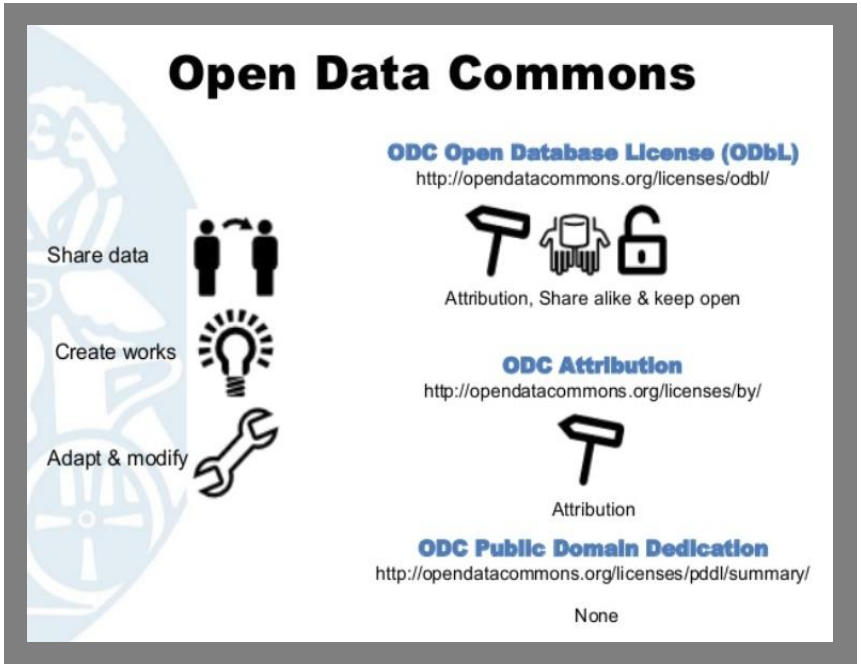
Is data transparent?

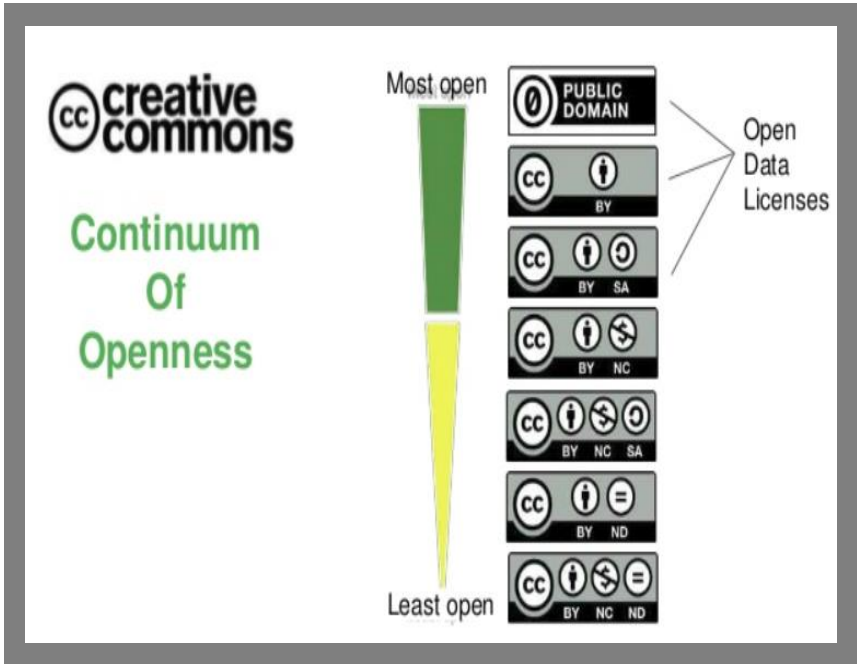
Meaningful insights are only possible with high quality data

- ❑ Is openly licensed?
- ❑ Is machine readable?
- ❑ Is free of charge?
- ❑ Is available in bulk?
- ❑ Is up-to-date?
- ❑ Is online?
- ❑ Is digital?
- ❑ Is available?
- ❑ Is publicly available?



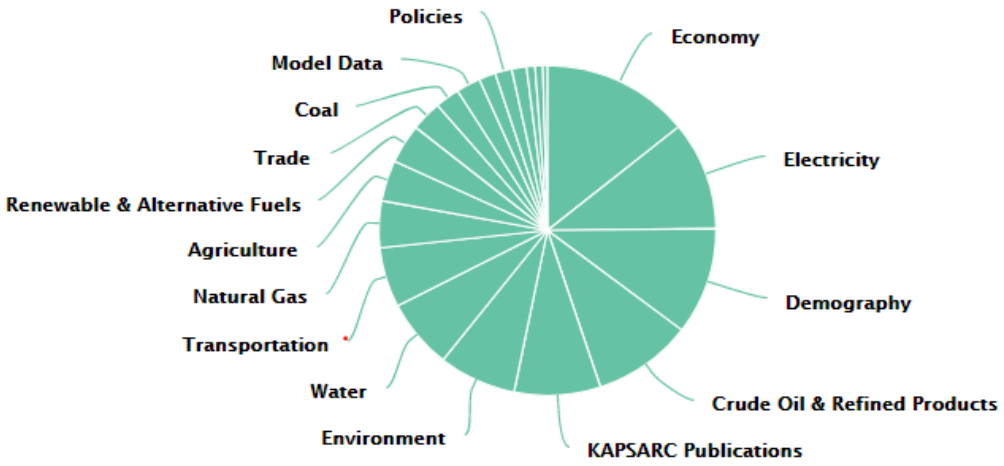
Data owners need to specify “data use” standards, soon technology will aid





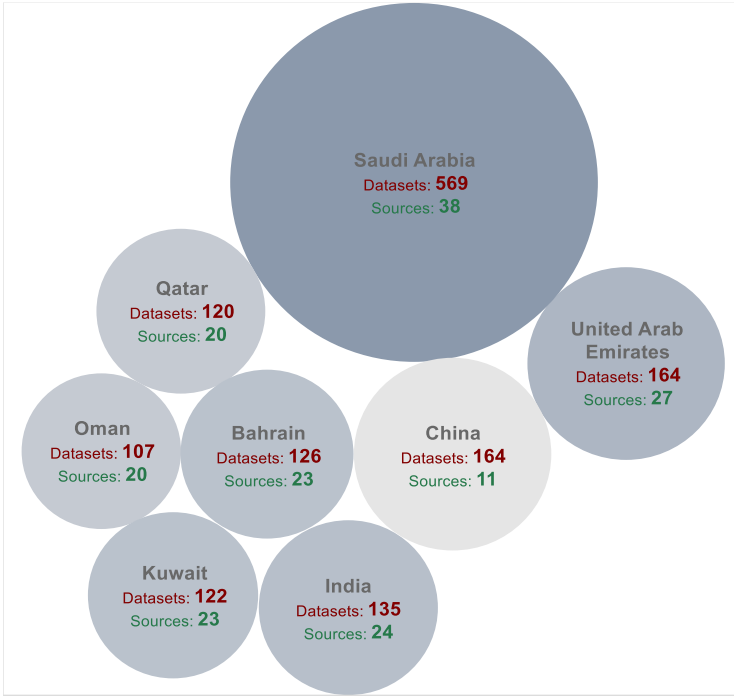
Datasets and themes

- **30 million records**
- **1700 Datasets**
  1000 public - 700 restricted
- **150 Sources**
  Identifying key sources from GCC, China and India
- **16 Themes**
  3 categories



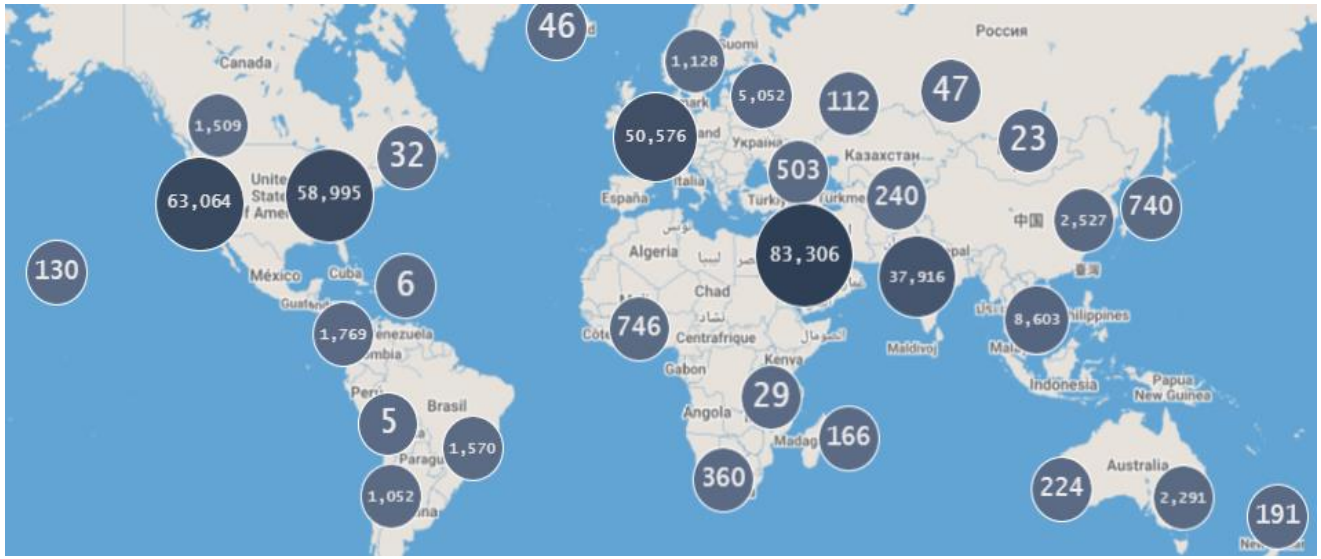
Top countries data coverage

- **GCC** 60+ data sources 1200+ datasets
  - **Saudi Arabia** 35+ sources
  - 560+ datasets
- **India and China** 35+ data sources 290+ datasets



Key take away

- Collaborate on granular data availability to advance research insights
- High frequency data, currency, machine readable data with auto alerts
- Extend data, model, insights via apps to improve understanding of energy economics



Let's collaborate on open

Data

Models

Tools

Insights