Towards AI and Data Commons

Towards AI and Data Commons

AI for GOOD
GLOBAL SUMMIT

- identify **practical applications of AI** with the potential to **accelerate progress** towards the **United Nations' Sustainable Development Goals**.

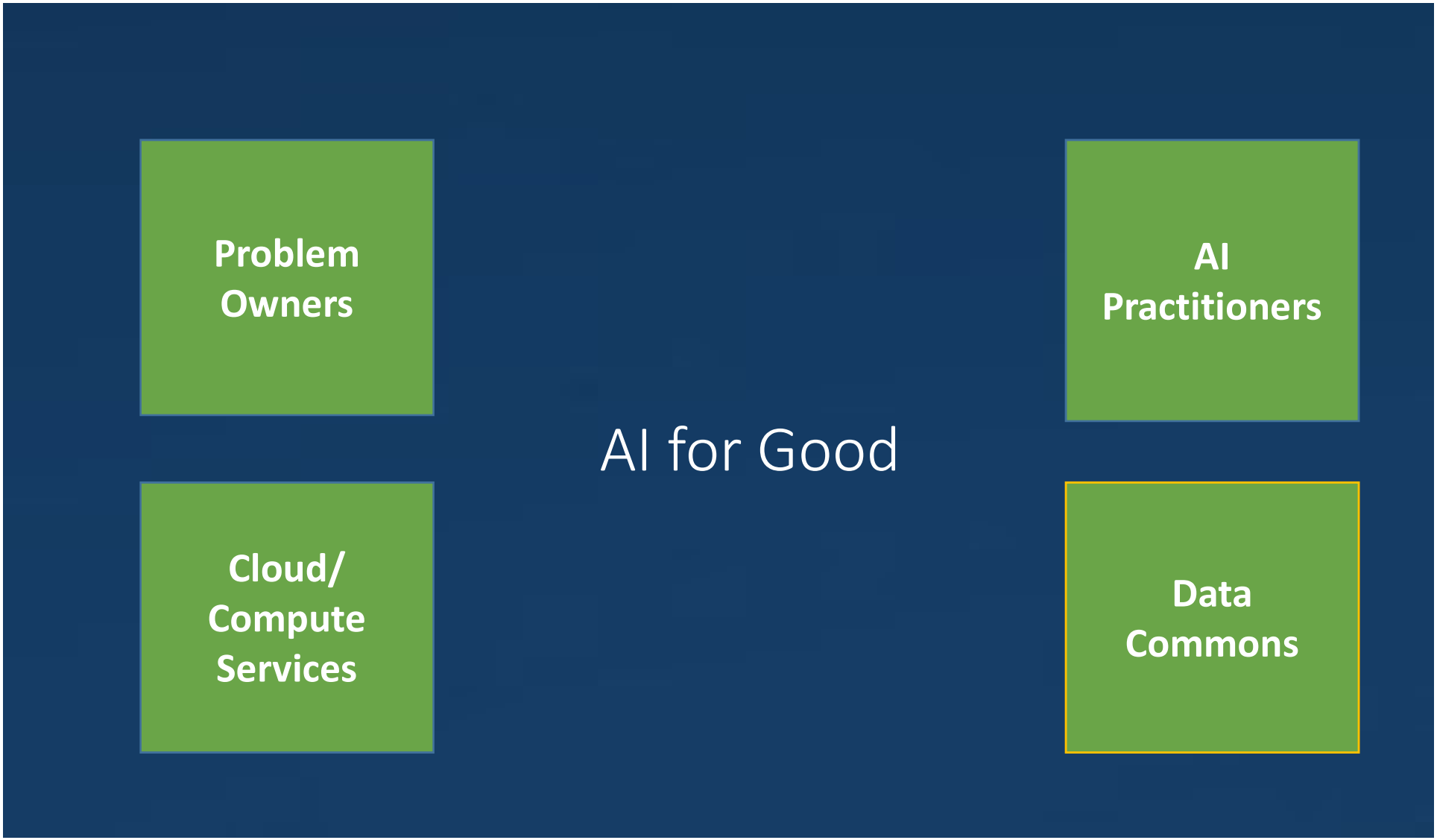
- formulate strategies to ensure **trusted**, **safe** and **inclusive** development and dissemination of AI technologies and **equitable** access to their benefits.

AI COMMONS

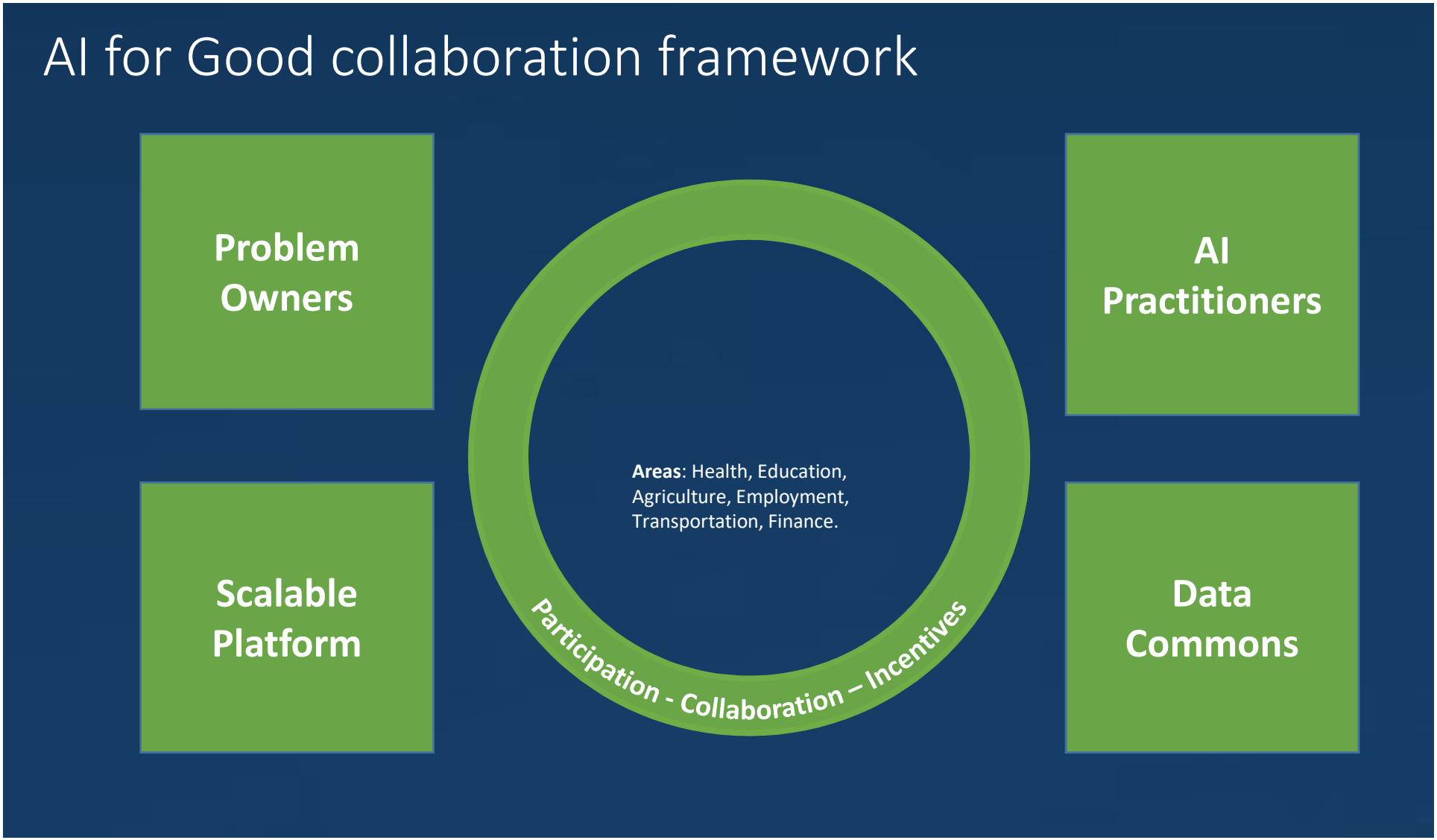
A Common Platform
for AI for Good

Opportunities

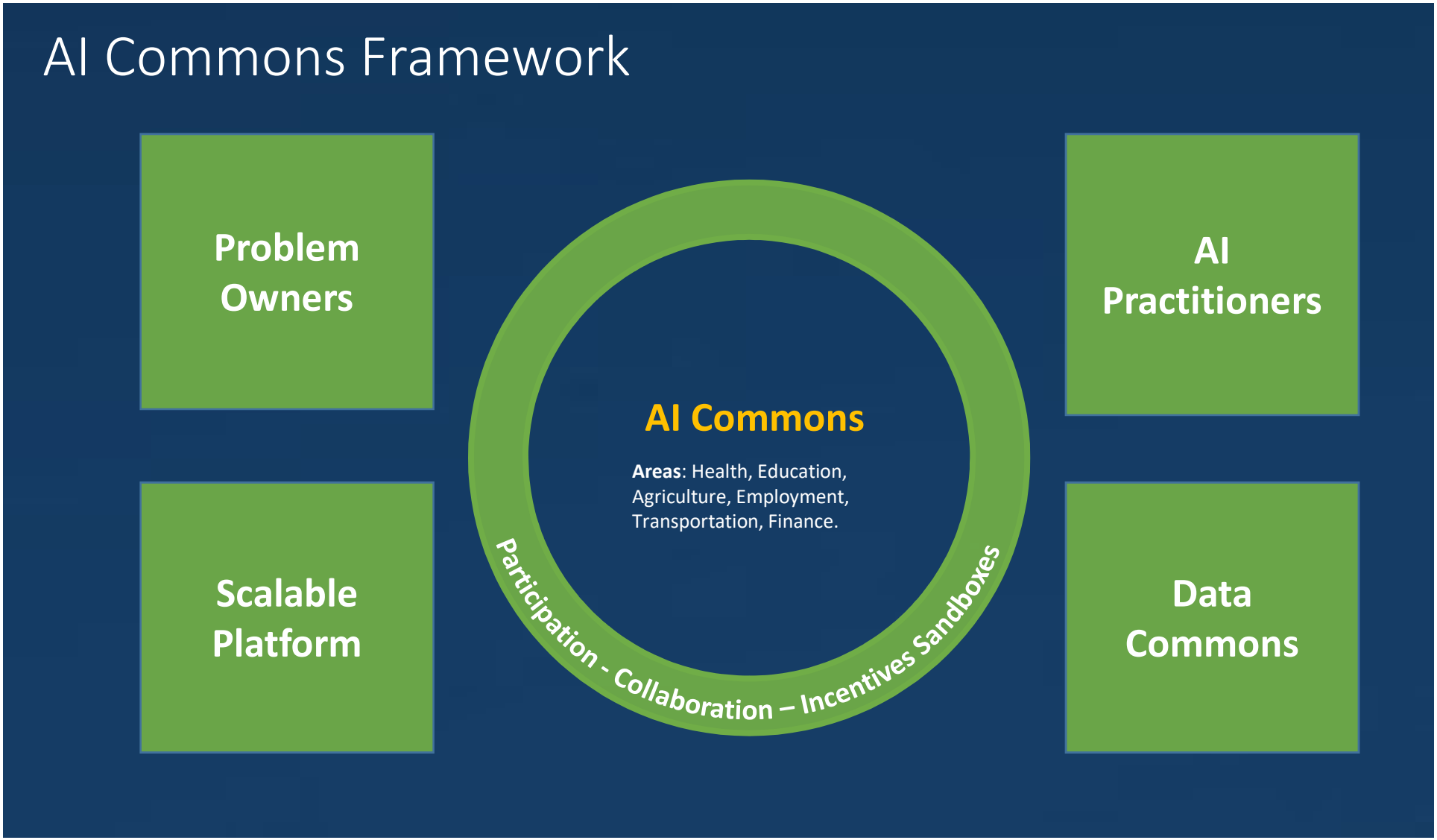
- To use AI to help solve many problems not currently addressed
- To identify core problems and make them visible to AI practitioners
- To facilitate collaboration between problem owners and AI practitioners
- To provide a “safe environment” to help evaluate feasibility of AI solutions
- To help transition pilots to global services


Problem Owners
AI Practitioners
AI for Good
Cloud/ Compute Services
Data Commons

AI for Good collaboration framework



AI Commons Framework



Example: Project 0 proposed from Satellite Track

Opportunity: Provide infrastructure platform to deliver continuous, permanent global services based on automated analysis of satellite++ data streams

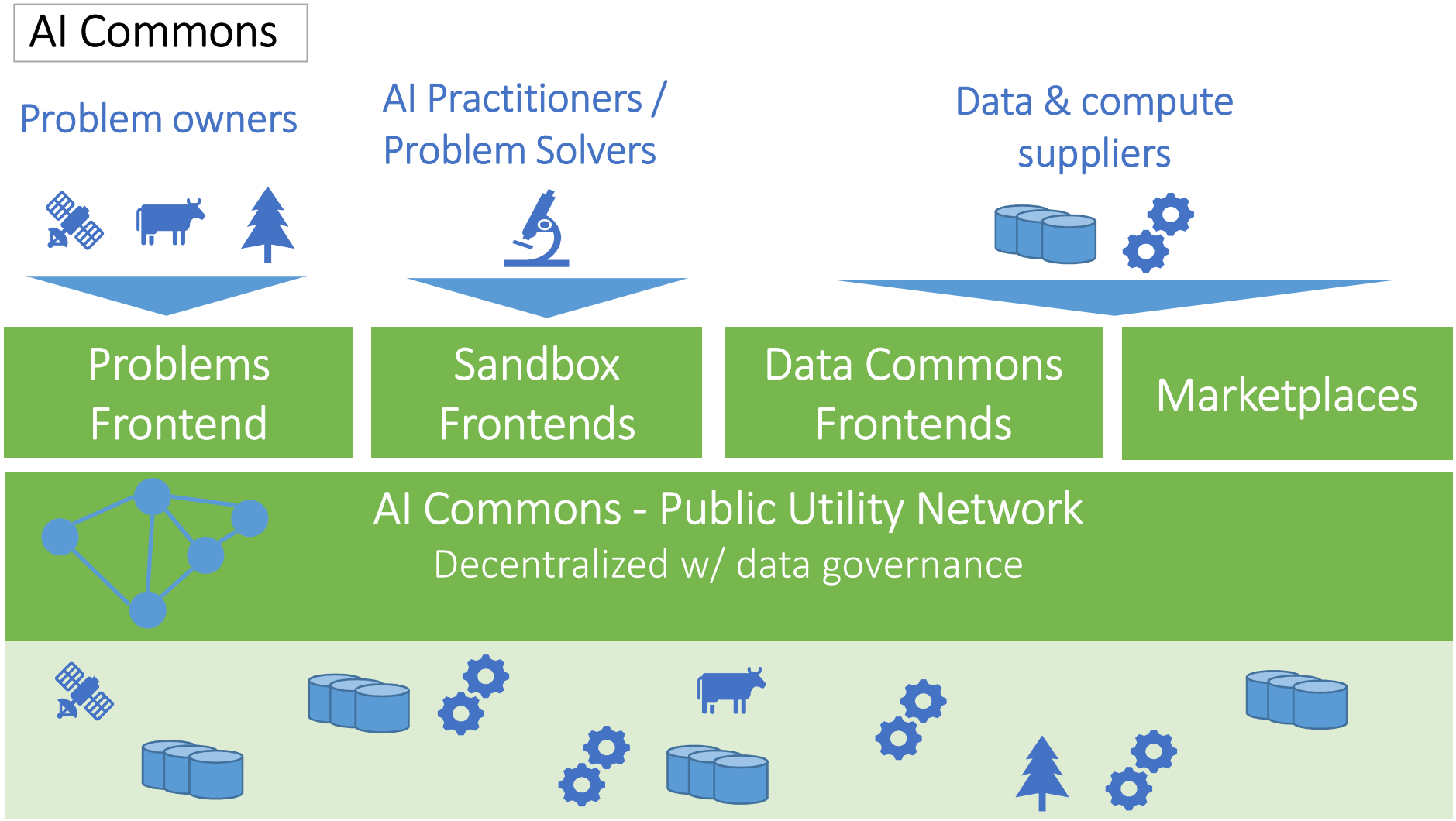
Examples of global services:

- weather data and forecasts
- urban traffic state
- global forest watch
- CTBT global seismic monitoring

Benefit of the AI Commons: for each application,

- focus only on the analytical capability and “user interface”
- transition from pilot to global service is immediate

A model for AI Commons


AI Commons
Problem owners
AI Practitioners / Problem Solvers
Data & compute suppliers
Problems Frontend
Sandbox Frontends
Data Commons Frontends
Marketplaces
AI Commons - Public Utility Network
Decentralized w/ data governance

Features of the platform

High Level

- Decentralized – no single entity owns or controls
- Free *and* priced – reconcile data commons & data marketplaces
- Variety of data – raw training data (structured & unstructured), cleaned data, models
- Built-in curation – high-quality datasets bubble up
- Incentives to share data – system rewards

Privacy Preserving

- Privacy preserving storage – local -> cloud -> decentr. cloud
- Privacy preserving compute – “”

Data Governance

- Global and local / sandboxes – permissions give flexibility in access to different datasets
- Digitally signed – securely know which actors did what
- Provenance – history of data & compute from origin onwards
- Labels – e.g. “space”, “binary-valued”. Emergent ontology

Legals & Interoperability

- Network of >1 utility networks
- Interoperable IP standard – COALA IP
- Template legal documents for licensing

An AI sandbox for all

FRANCESCA ROSSI

IBM RESEARCH AND UNIVERSITY OF PADOVA

AI for all and from all

- AI will impact all ➔ all need to have a voice in AI
  - And means to express that voice
- AI ingredients made available and simple to use
  - mix and match approach
  - compositionality
- Very practical framework, tangible and concrete collaboration mechanisms
- Collective decision making, across borders
- Not just data commons
  - AI is not just data-driven: ML, but also planning, scheduling, search, KRR, etc.

Who should be involved

- Not just AI pratictioners and problem owners
- Also AI researchers, social scientists, data subjects, policy makers
- Only approach to make it successful:
  - Multi-cultural
  - Multi-gender
  - Multi-disciplinary
  - Multi-stakeholder

Trustworthy AI for all and from all

- Fairness
  - Bias detection, mitigation and auditing tools
- Value alignment
  - Community, sector, and task dependent
- Explainability
  - To build trust over time
- Also tools for designers and developers to help them think about ethical issues and resolve them
- Collaboration/partnership with existing initiatives around AI ethics and trust
  - PAI, IEEE EAD, TrustFactory.ai, etc.