

# Big Data, Smart Cities, I4.0: Trends & Threats

June 18 2019

Bharadwaj Amrutur  
Professor, Indian Institute of Science

# Harvesting value from data

## Trends

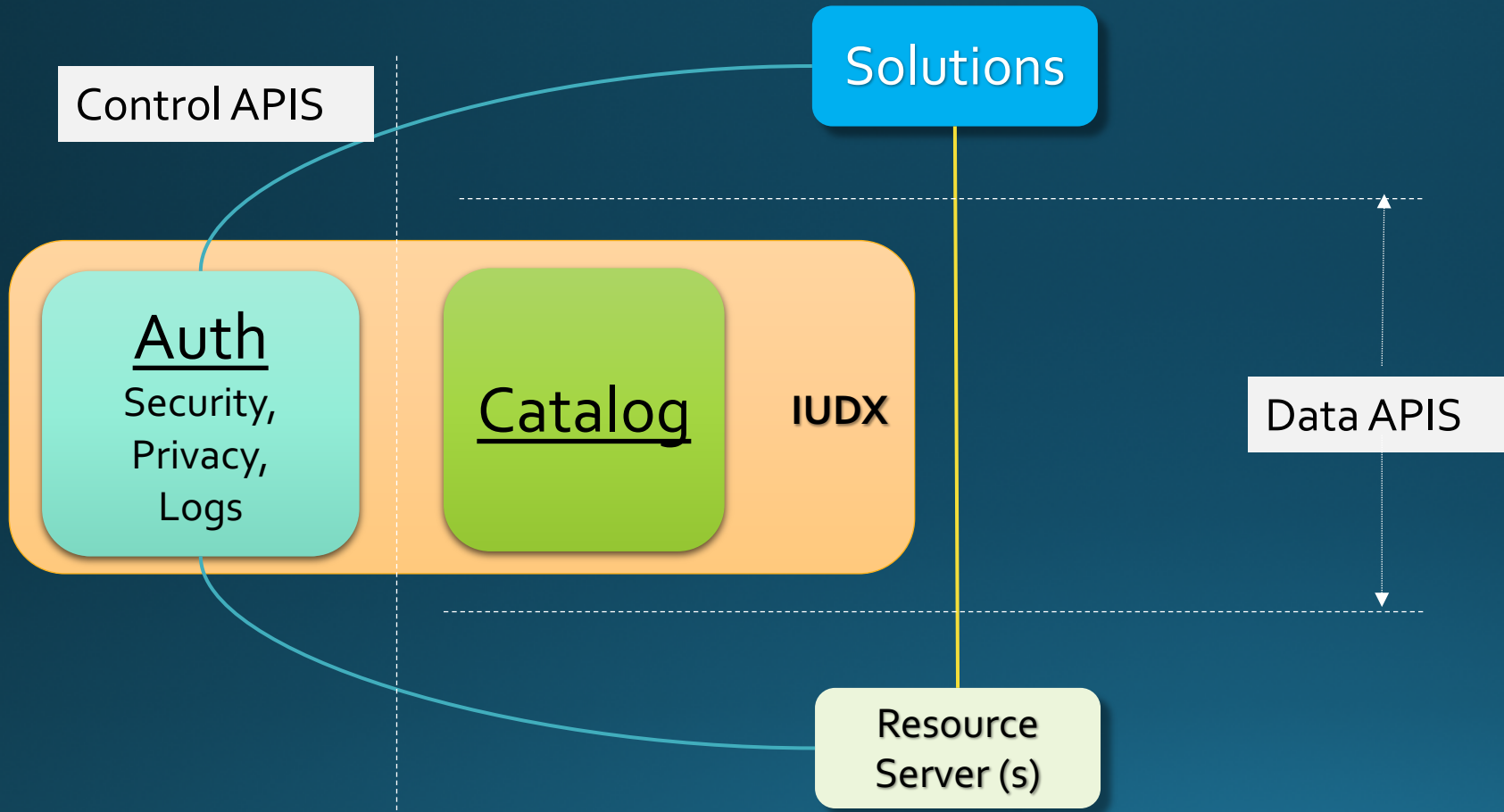
- Data generation/creation has become easier
- Powerful open source tools and frameworks for data analytics
- Convergence of government and enterprise (in terms of ICT)

## Threats

- Friction for data flow across silos
- Lack of standardized ways for discovering and accessing data
- Poor data quality
- Weak privacy and security
- No sustainable business model
- Ignorance of best practices

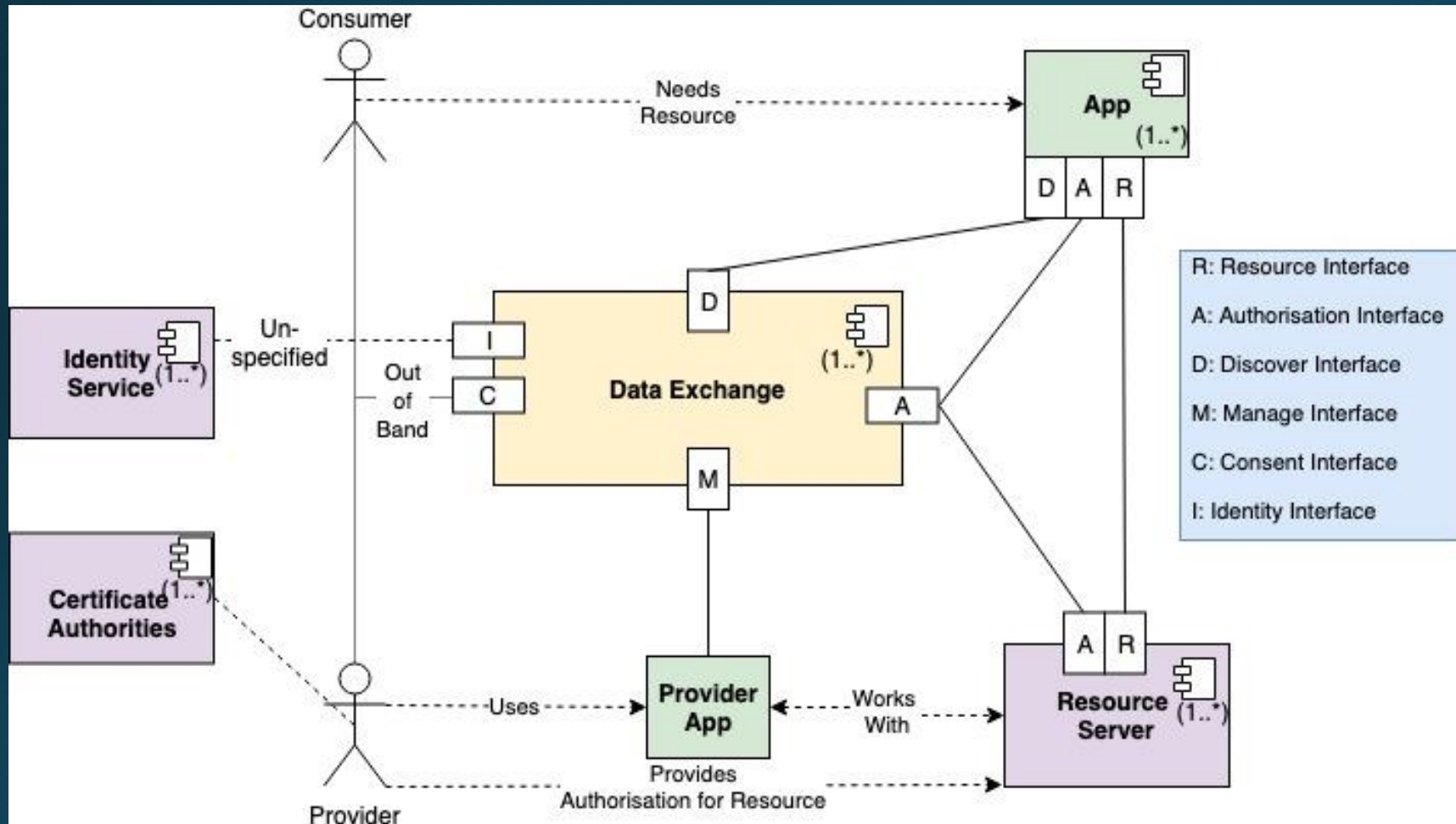


# India Urban Data Exchange: (IUDX)



- Authorization:
  - Existing: Authorization within organization
  - IUDX: Extends across organizations.
  - Data Privacy by Design
- Catalogue
  - Stores meta-data
  - Enable discovery of data.
  - IUDX: Supports data taxonomy (JSON-LD) and type checking (JSON-Schema)
- Data APIs
  - Work with existing standards (and proprietary)

# IUDX Authorization



- X.509 Certificate based authentication
- Enable data custodians to manage sharing settings and policies
- Federated authorization based on industry standard Oauth/UMA2.0

# IUDX: Data Privacy

Labels → Meaning.   v	Public	Protected	Private	Confidential
<b>Nature of data</b>	Contains no personal information	Contains anonymized information	May contain personally identifiable information	May contain personally identifiable information and/or other data that is confidential within the organization
<b>Authorization protocol and policy</b>	None	Requires authorization using UMA, no custom auth policy	Requires authorization using UMA, custom auth policy specified in XACML	Requires authorization using UMA, custom auth policy specified in XACML
<b>Consent</b>	None	Provider	Requires consent of owners	Requires consent of owners
<b>Data locality</b>	None	None	Configurable or as per regulatory framework	Only service based access
<b>Data retention</b>	None	None	Configurable or as per regulatory framework	NA
<b>Data storage</b>	Any	Any	Encrypted	NA
<b>Data usage</b>	Any	License	Licensed with legal framework	Licensed with legal framework
<b>Data audit</b>	None	Random audit	Needs audit	Needs audit
<b>Data Monetization</b>	Not to be monetized	Provider's decision	Provider's decision	NA
<b>Data Processing Enclave</b>	None	Provider's decision	Provider's decision	Required

# IUDX Catalogue - 1

- Enable discovery of datasets based on queries
  - E.g. What datasets are available related to water and women's safety in Khan Market's neighborhood?
- Meta information about the datasets so that
  - Consumer knows whom to connect with to get permission
  - Description of the data model (what are the fields, what they mean etc)
  - Possible linkage to a more detailed taxonomy or even an ontology
    - Taxonomy : is a dictionary which explains the terms
    - Ontology: is a dictionary along with logic inference rules to allow machines to do reasoning
  - API Information about how to access the data: Needed for a developer



# IUDX Catalogue – 2: Concepts

@context: <http://iudx.org.in/safety-taxonomy>

Publisher: safetipin.com

Tags: [women's safety, safety index, urban quality]

IsPartOf: NirbhayaScheme

Datamodel :

Lighting : {type: integer, description: "amount of lighting, 0-5"}

Feeling: {type: integer, description: "feeling of safety, 0-5"}

Location: {type: GeoJSON, description: "location of data"}

SafetyIndex: {type: integer, description: "safety score 0-5"}

ImageLink: {type: URL, description: "link to access raw image"}

API :

apiLink: <https://swagger.api.com/safety>

Link to external  
vocabulary or ontology

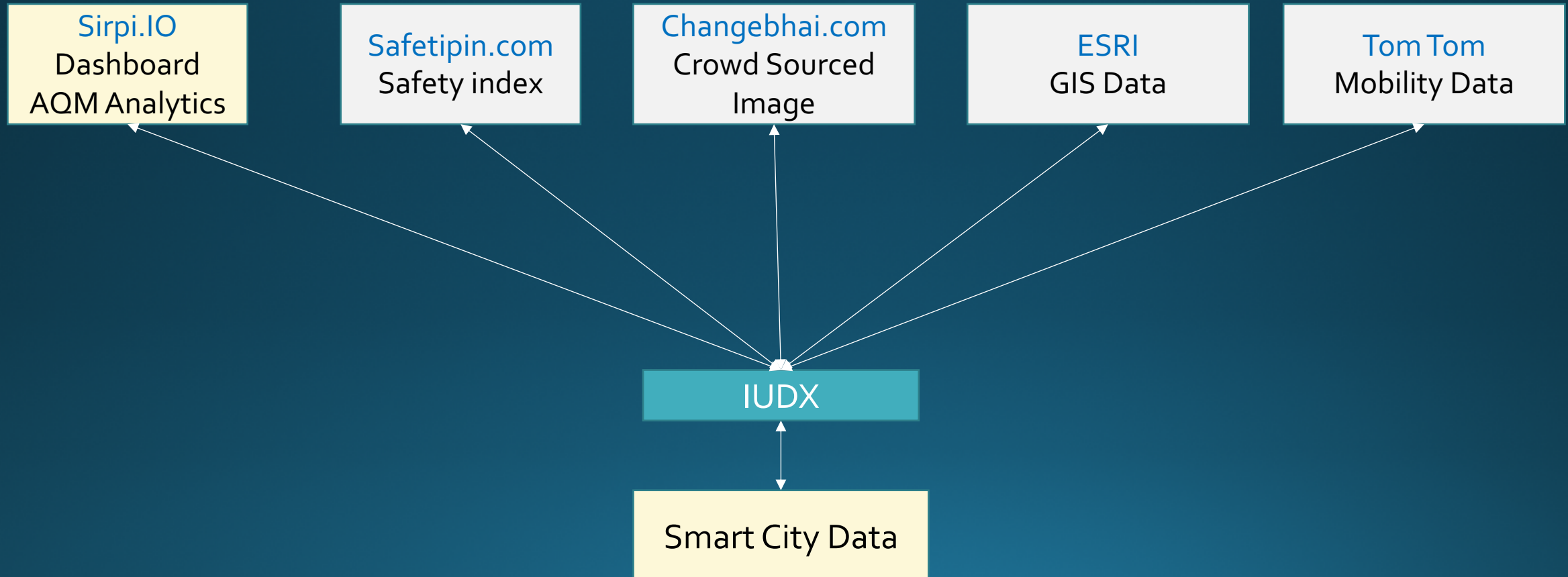
Aid search/discovery

Define relationships

Basic documentation

API Documentation

# Pune IUDX Ecosystem (pilot in progress)



# Summary

- Reduce friction to enable data to flow across silos
  - Privacy and Security
  - Data discoverability
- Create market and business ecosystem
  - Incentivizes exposing of data
  - Incentivizes good data quality