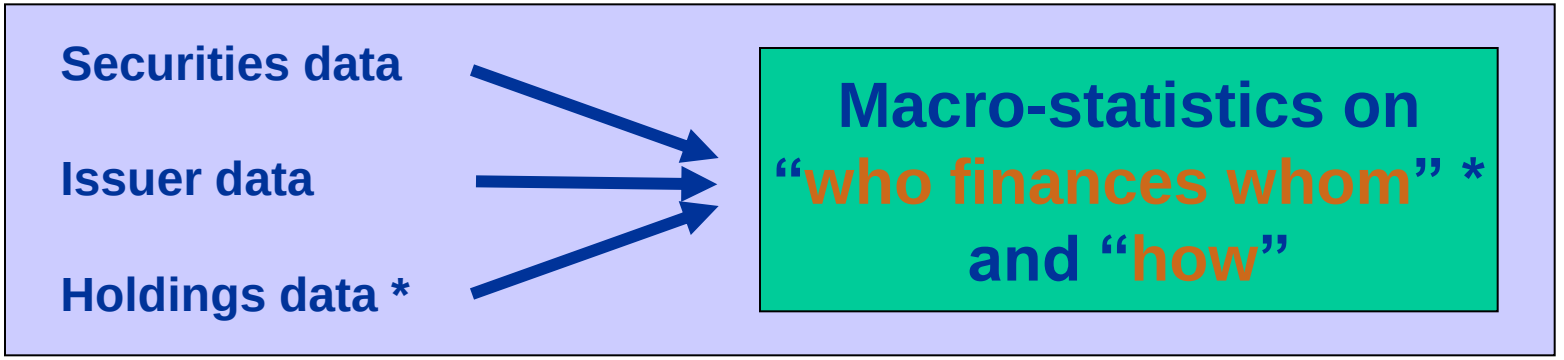
MIT Information Quality Industry Symposium
15 July 2010

Reference Data Utility

Francis Gross
DG Statistics, External Statistics Division

Financial statistics based on micro data



* aggregated by economic sector and country of residence – under development

- **Benefits**
  - Flexibility in serving event-driven policy needs, near time
  - Ability to drill down, linking macro- to micro-issues

- **First tool: the Centralised Securities Database (CSDB)**
  - Holds data on nearly 7 million securities
  - Is in production with 27 National Central Banks online

CSDB Data Quality Management made us face a choice:

either

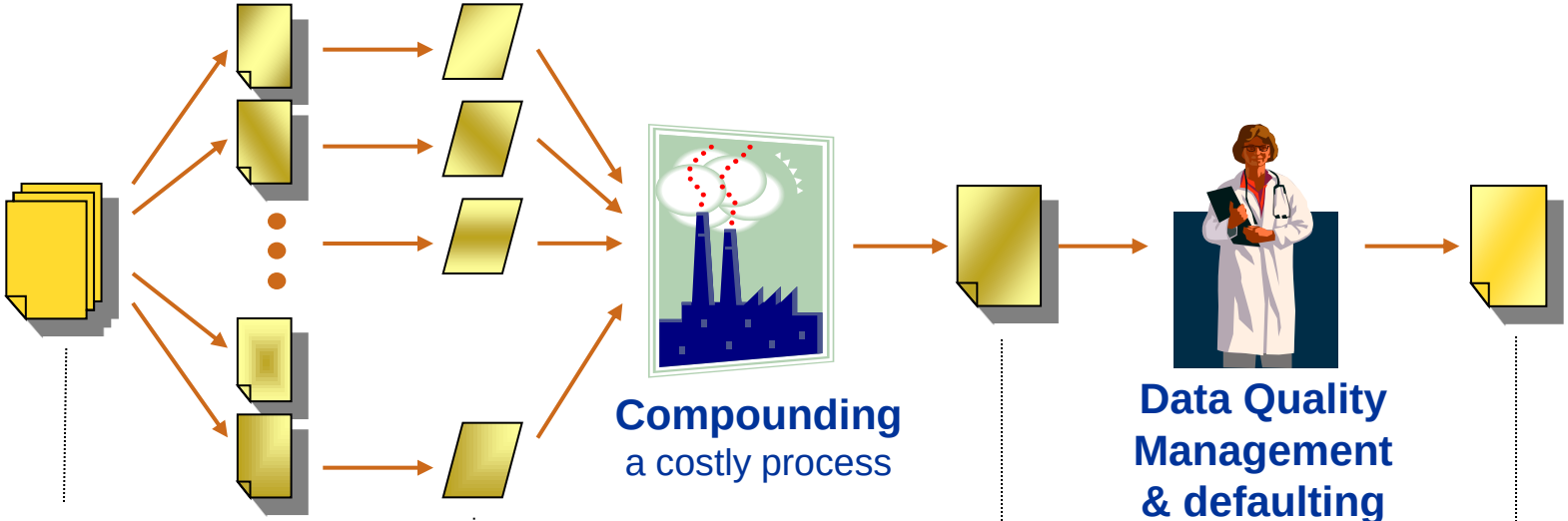
SYSYPHUS

or

CHANGE,

Change beyond Statistics

The CSDB data supply chain



Prospectus
"perfect", public data source, but no common language

Commercial data sets
error prone, selective, costly production, duplicate efforts, proprietary formats

Candidates
for compounding; a costly coll

Compounding
a costly process

Compound
first shot

Data Quality Management & defaulting
Costly, burdensome, often just palliative

Golden Copy
after DQM & defaulting, still not back to perfect - and not standard.

Duplication and non-standardisation in the very data capture process hamper the whole downstream value chain.

Tectonic plates have moved, slowly.

Tensions have built, imperceptibly.

60 years of progress changed the context

- Progress in politics, education, technology
- made markets global, more complex and faster,
- hitting human limitations.
- No choice: we must trust automation.
- IT can do that, but data is now a major bottleneck;
- Ever more independent sources: each one a "data dialect".
- Even if differences are subtle, this defeats IT and creates costs.
- For all of us alike, industry and authorities.

The Tower of Babel, again... This time with data.

The crisis showed: we need to act now!

The crisis put data on centre stage

- Weak data certainly didn‘t help in the current crisis.
- Promises were made to the people:
  - Markets will be made more transparent;
  - Markets will be monitored for systemic risk.
- Technical pre-requisites – necessary conditions:
  - Capability to run industrial-, large-scale computing on
  - Vast pools of high-quality, up-to-date micro-data; and
  - Capability of fast, ad hoc data collection when crisis erupts.
- Industry needs better data as well, for efficiency and op^{tl} risk.

We all need the same good basic reference data.

Why build more than one infrastructure?

Shared reference data infrastructure: a must

- Data will be good for all users or for none.
- Diversity in "data dialects" must be reduced;
- For that we need true standards, i.e. discipline and rigour.
- Reference data on instruments & entities: a good place to start
- Standardised data infrastructure: a shared strategic resource.
- The financial industry did so far not manage by itself:
- Top down action seems needed to converge in a finite time.

A shared reference data infrastructure is needed.

Legal compulsion seems necessary.

Where to start?

The first layer of data captured from reality.

Its generation process matters

Data capture drives IT output quality

- Once good data is in the system, processing can work well.
- Data capture from the "real" world is the key step.
- Once lost at capture, information in data is lost:
- No "data cleaning" will help: data must be captured again.
- Messy data capture at source is very expensive downstream:
  - Most applications perform badly
  - "Data cleaning" and fixing failed processes are costly for all
  - Processes and IT must be designed in complicated ways

Large scale IT processing can be simple and cheap when data fulfils the programmers' quality assumptions.

Messy data capture delivers "garbage in, garbage out".

Progress is on its way

“…a standard for reference data on securities and issuers, with the aim of making such data available to policy-makers, regulators and the financial industry through an international public infrastructure.” (J.C. Trichet, 23.2.09)

Dialogue with industry yields building blocks

- The EDM Council / IBM PoC tool for ABS and underlying
- Vendors begin developing Utility-like tools for local markets
- Cooperative models are springing up
- The EDM Council Semantics Repository: a major advance
- The entity identifier discussion is taking off again (ISO on BIC)
- The EDM Council / Carnegie Mellon “Data Management Maturity Model”: a data-focused management tool for the future

A viable reference data infrastructure benefits from constructive dialogue.

The industry expresses demand for a Utility

- **Industry panel at Conference 15 Feb 10 in London:**
  - **"An international Utility for reference data has its place, but**
  - **Keep it simple, *(concept of a "Thin Utility")***
  - **Ask industry to design the standards *(ISO does exactly that)* and**
  - **Give us the legal stick"**

A viable reference data infrastructure benefits from constructive dialogue.

The USA is just passing law

- **Office of Financial Research: a truly historic step !!!**
- **It is a first of its kind; there is no blueprint**
- **It is likely to become a blueprint for others**
- **There will have to be learning while doing**
- **It will have to deliver fast while building for the long term**
- **It will need to also develop a new organisation !!**
- **The international dimension is important (see Tarullo)**

Design concept and development process will determine lasting success

Data:

from

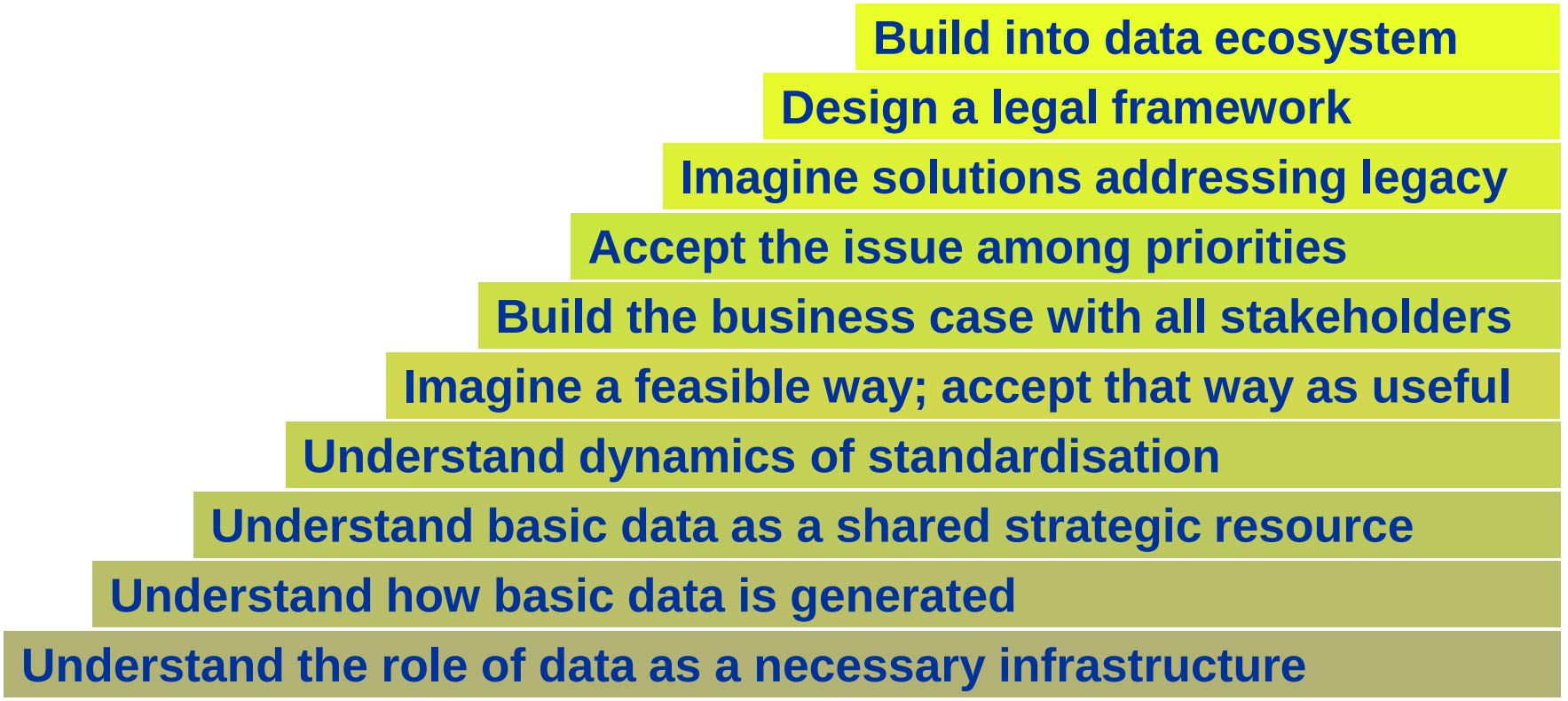
browsing and scavenging

to

farming

The long way to standardisation

Climbing the stairway to action



Business leaders, Policy makers, Regulators & Legislators now embrace the dialogue with the Data Community

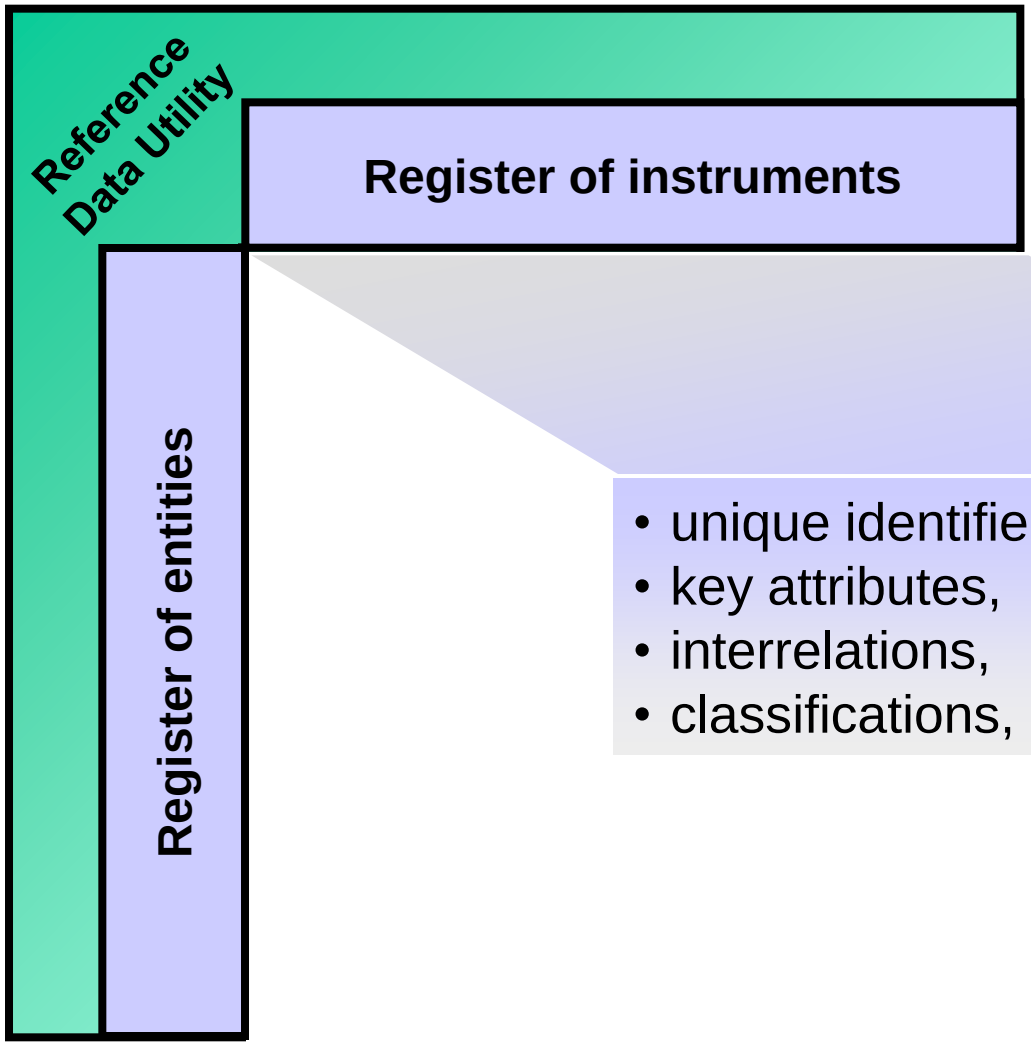
“Thin” Utility

“Thin Utility”: a unique, shared reference frame

- Two registers: one for instruments, one for entities
- Simple and light, complete and unequivocal
- Hard focus on identification and minimal description
- The shared infrastructure of basic reference data for:
  - Data users in the financial industry
  - Data vendors
  - Authorities
  - The Public
- An internationally shared infrastructure of reference

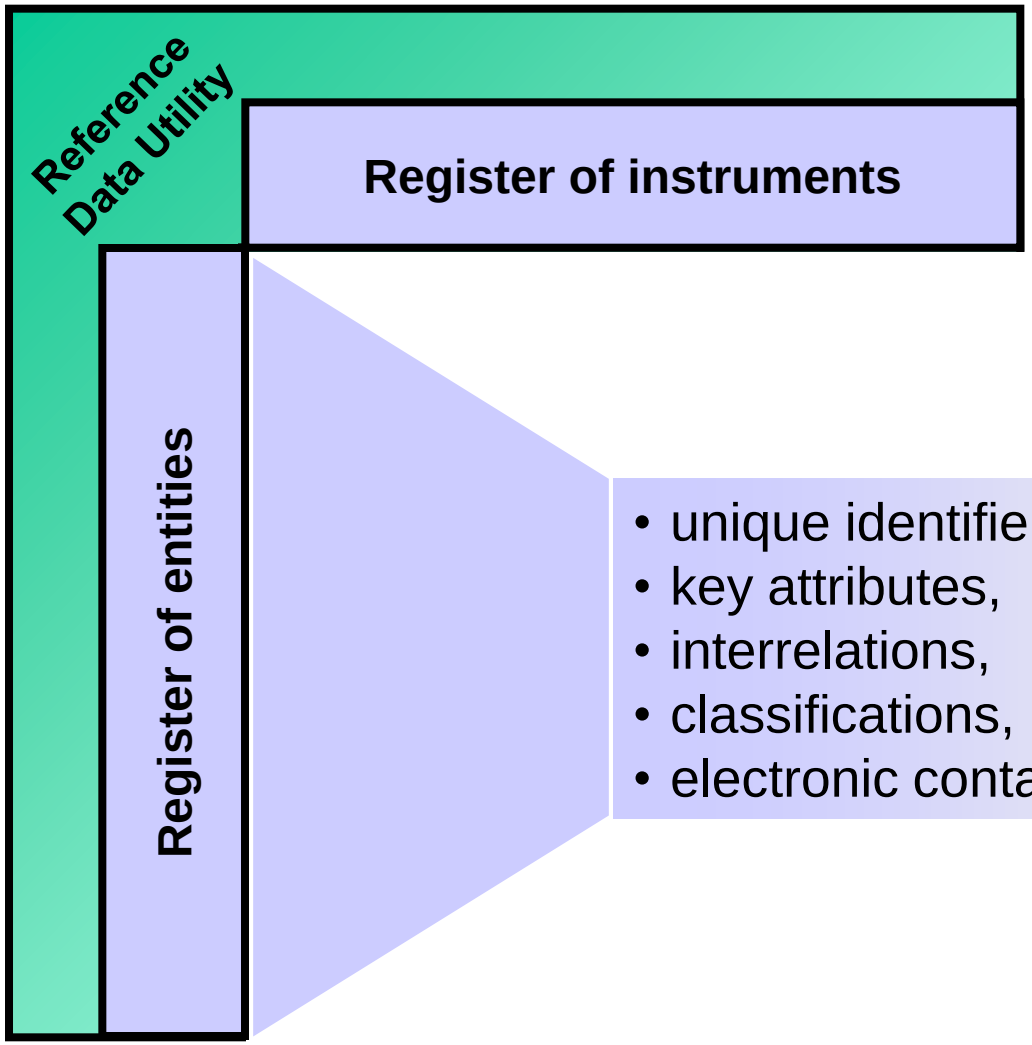
A “Thin Utility” provides the certainty of a single source on known, bare basics.

Two reference registers: the Thin Utility's frame

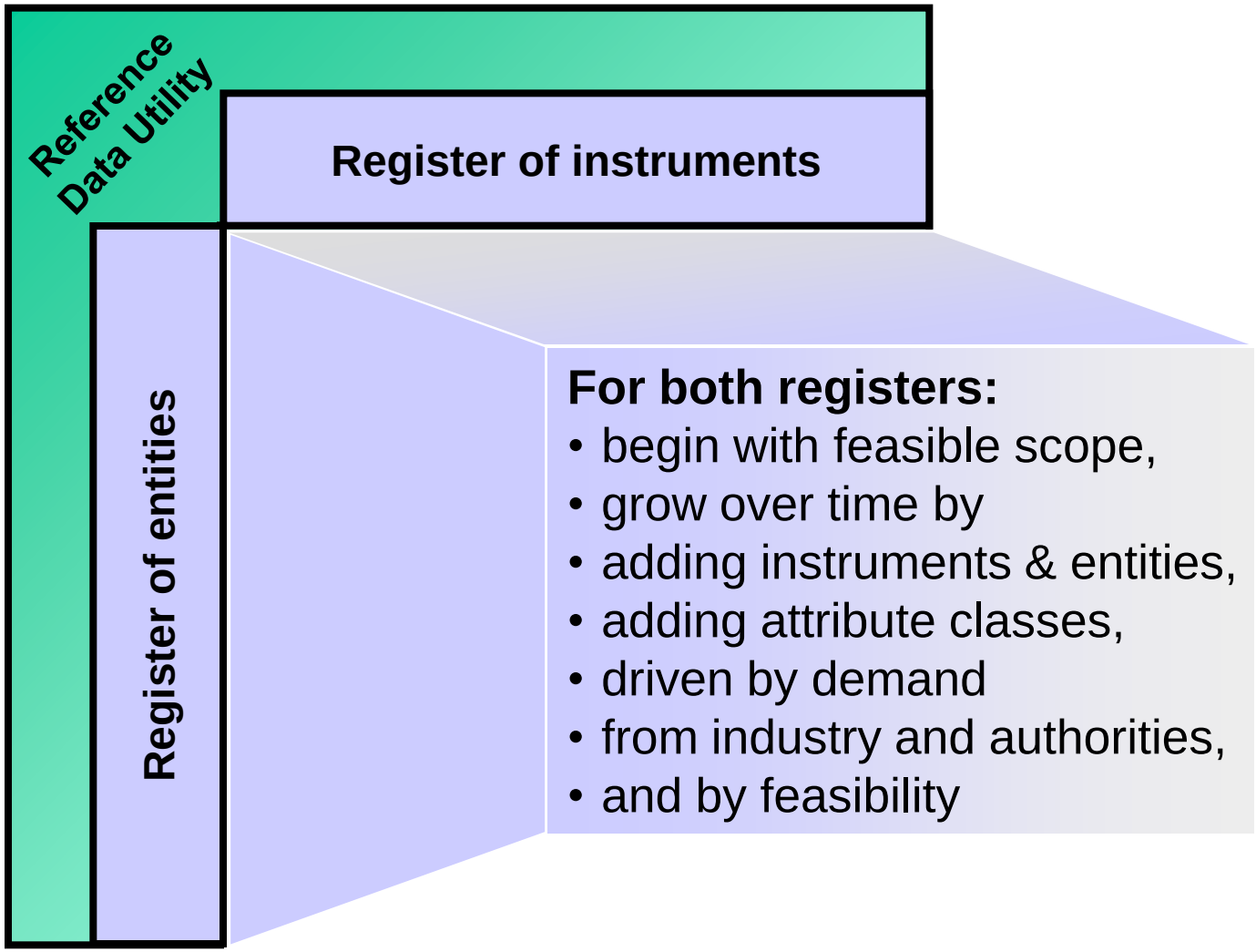


- unique identifier,
- key attributes,
- interrelations,
- classifications,

Two reference registers: the Thin Utility's frame

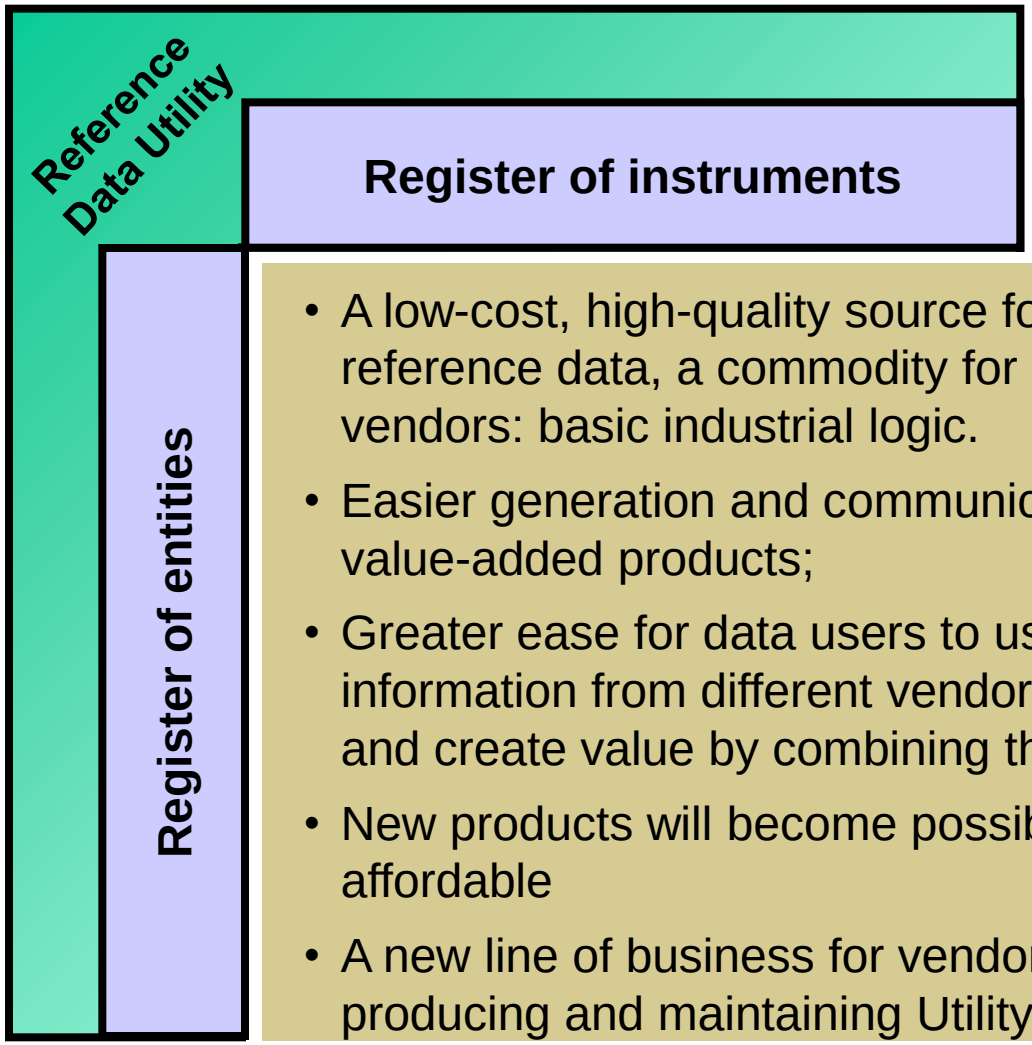


The Utility grows from a quickly feasible base



Supporting data vendors and users.

The Utility benefits data vendors & users



- A low-cost, high-quality source for core reference data, a commodity for all data vendors: basic industrial logic.
- Easier generation and communication of value-added products;
- Greater ease for data users to use information from different vendor sources and create value by combining them;
- New products will become possible and affordable
- A new line of business for vendors, producing and maintaining Utility data.

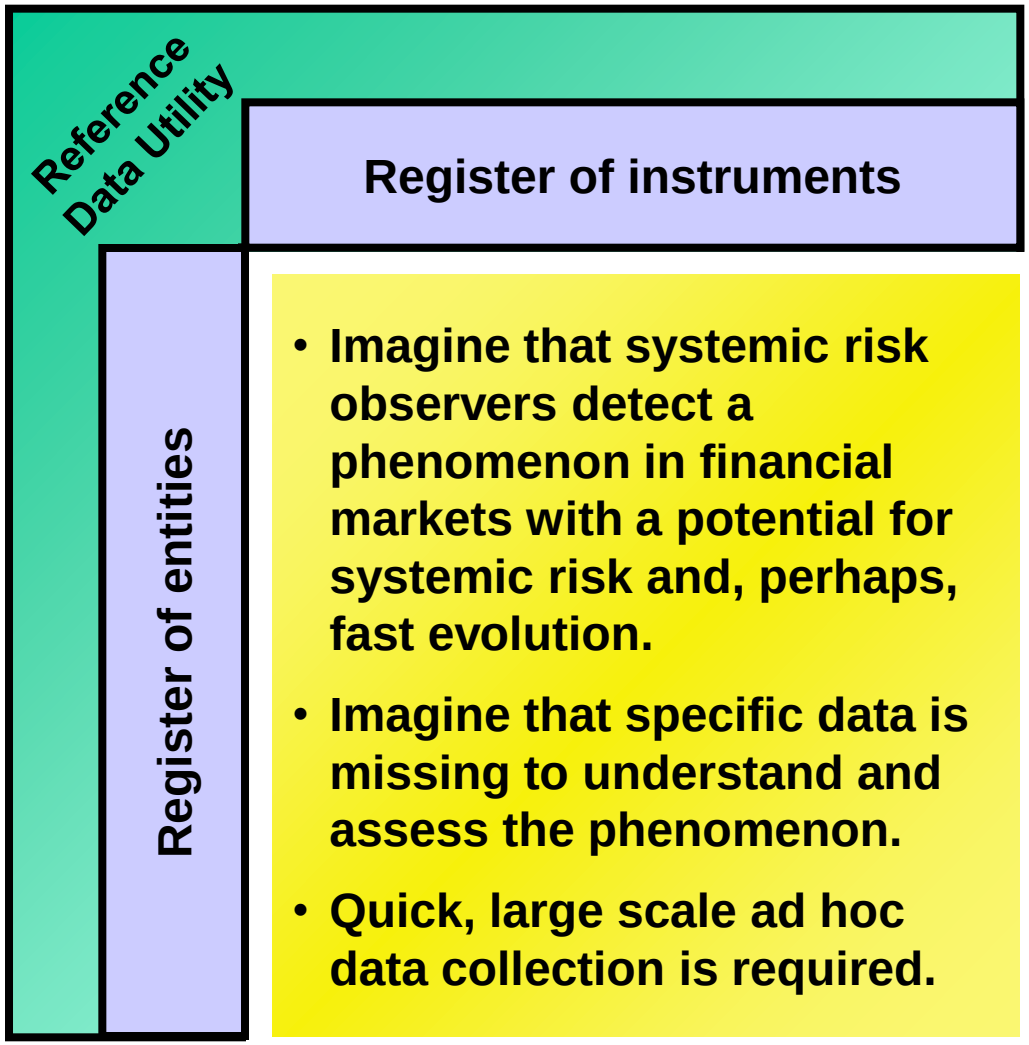
Supporting systemic risk analysis

Data challenge in the face of systemic risk

- **The economy, like any complex system, will surprise us: the next big crisis will come with surprising features.**
- **We will likely not have all data required at outbreak;**
- **We will thus need the capability to collect ad hoc data:**
  - **quickly,**
  - **on a large scale and**
  - **at a high level of quality and standardisation**
- **To enable fast analysis through large-scale computing, to**
- **Assess risk, and to test and dimension policy responses**

The technical capability to collect missing data fast must be ready when risk appears or when crisis erupts.

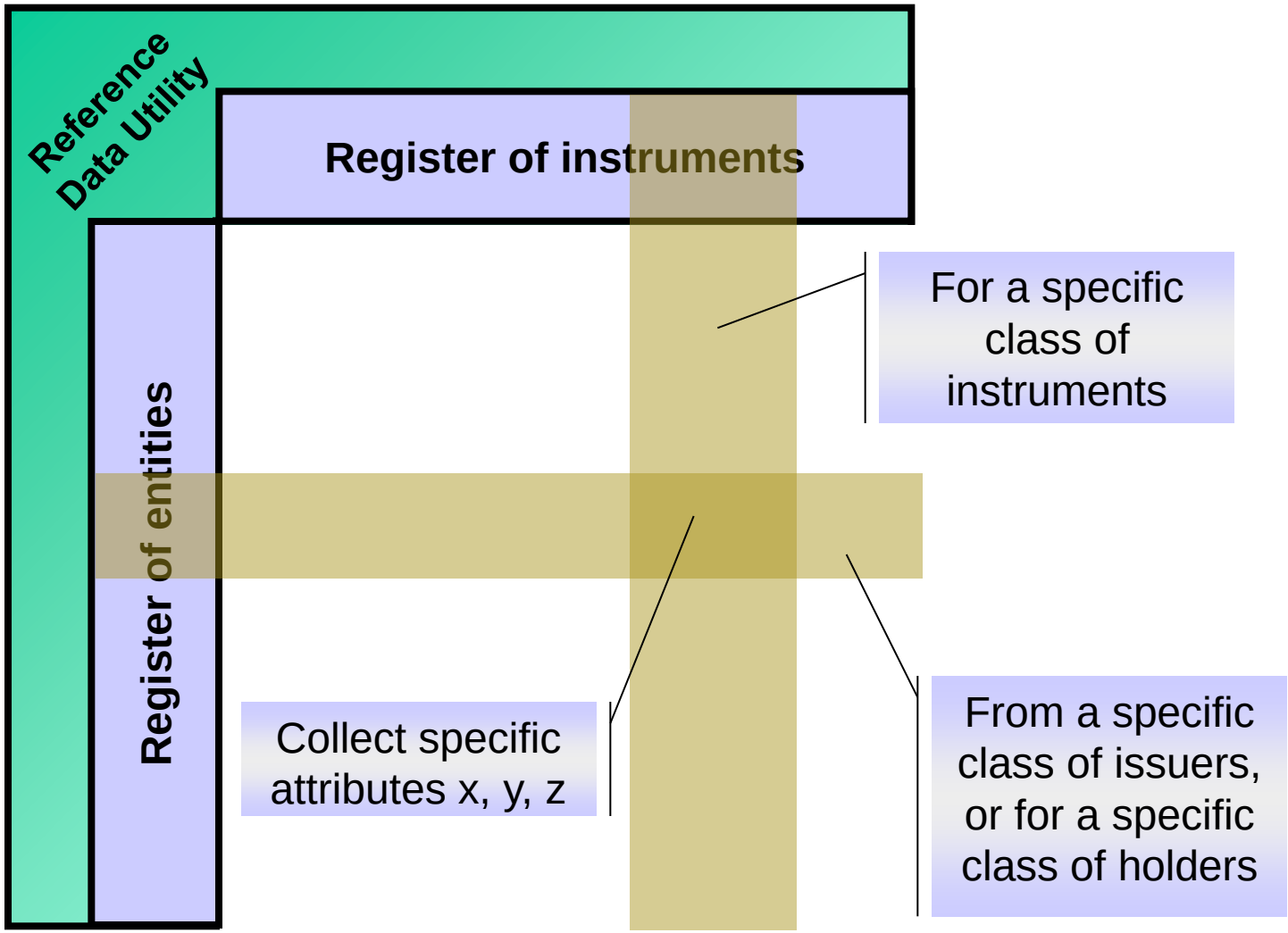
The Utility supports systemic risk analysis



- Imagine that systemic risk observers detect a phenomenon in financial markets with a potential for systemic risk and, perhaps, fast evolution.
- Imagine that specific data is missing to understand and assess the phenomenon.
- Quick, large scale ad hoc data collection is required.

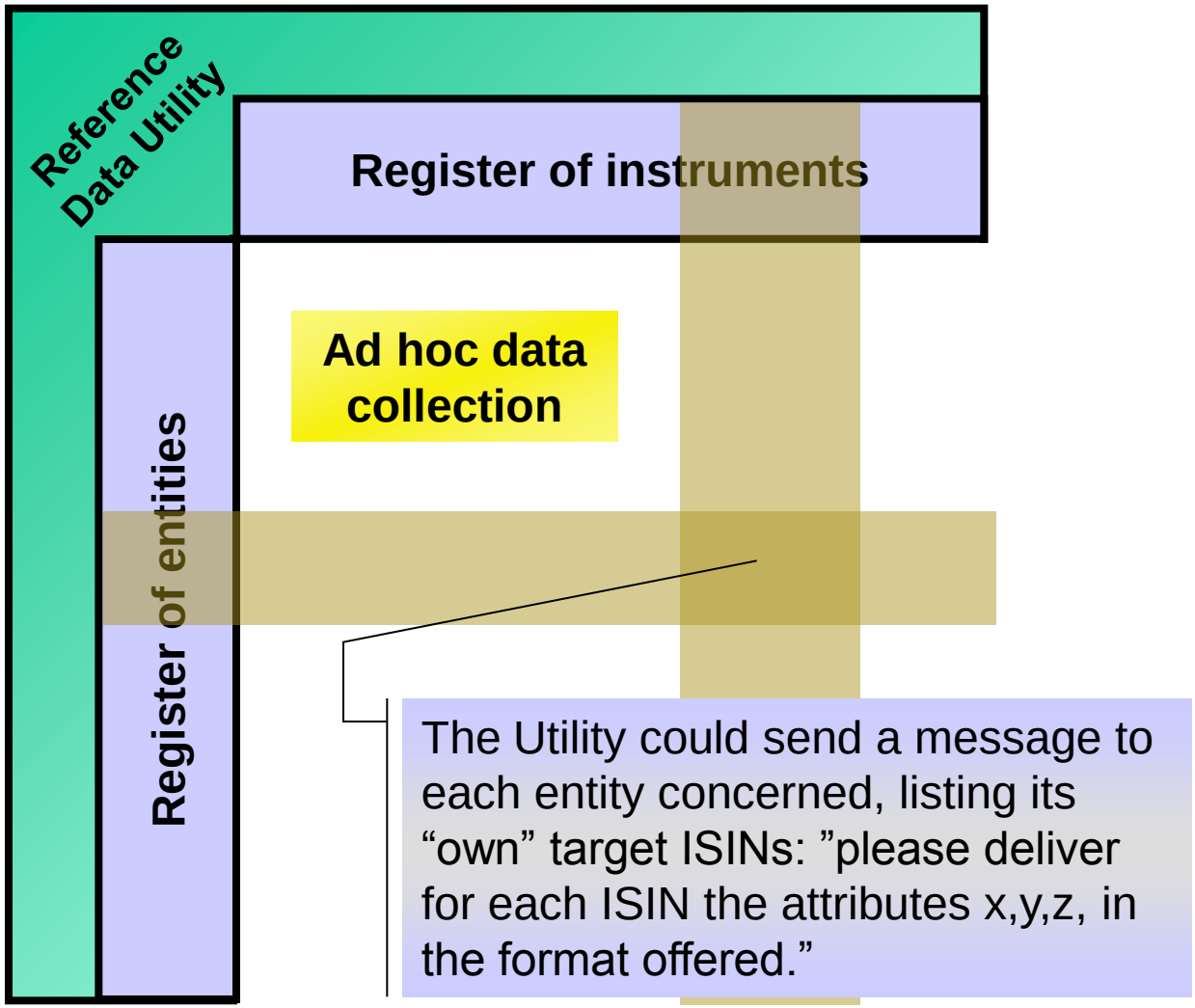
The Utility supports systemic risk analysis

the Utility supports ad hoc data collection



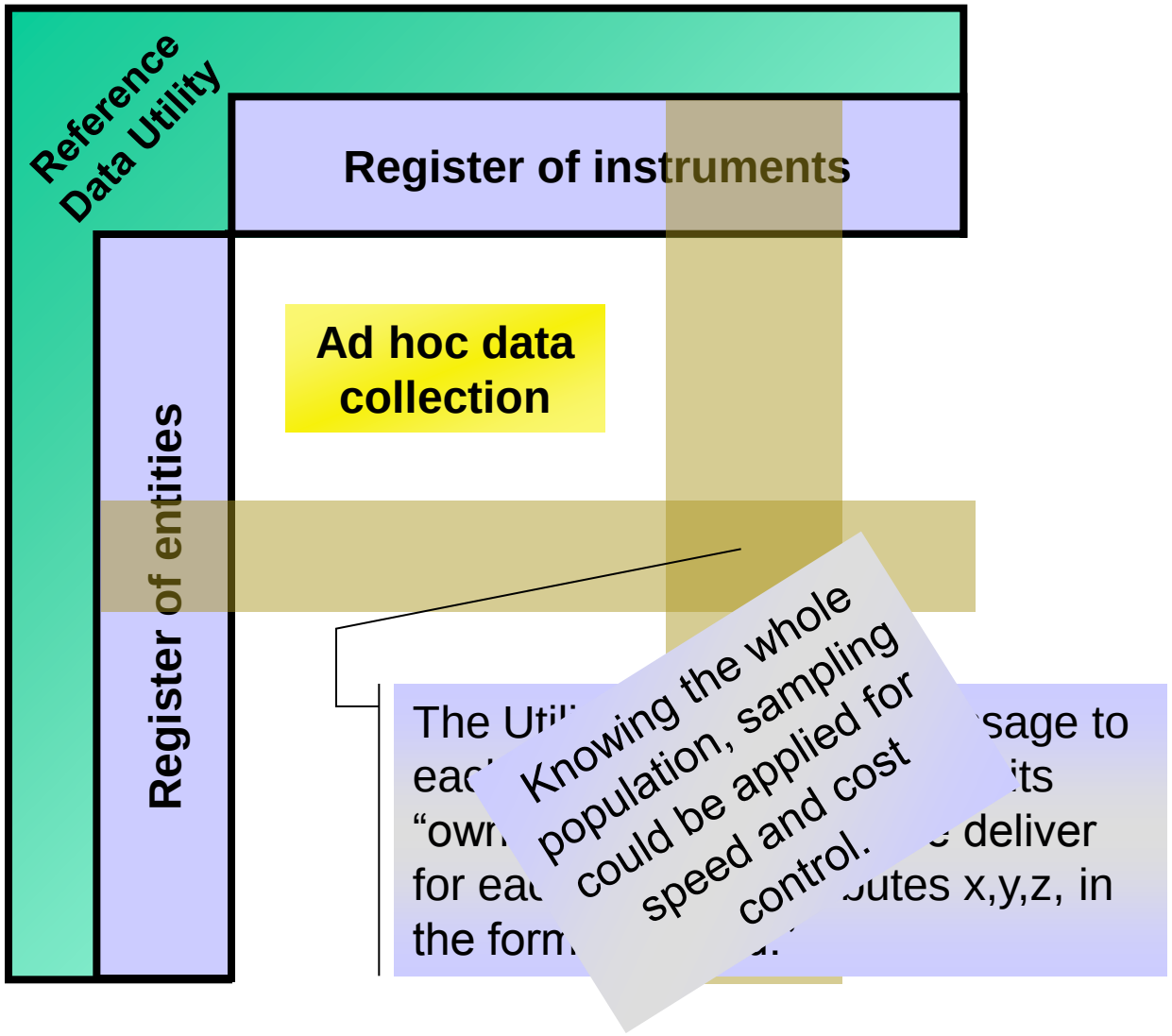
The Utility supports systemic risk analysis

A query from the Utility to many entities could take minutes only to program



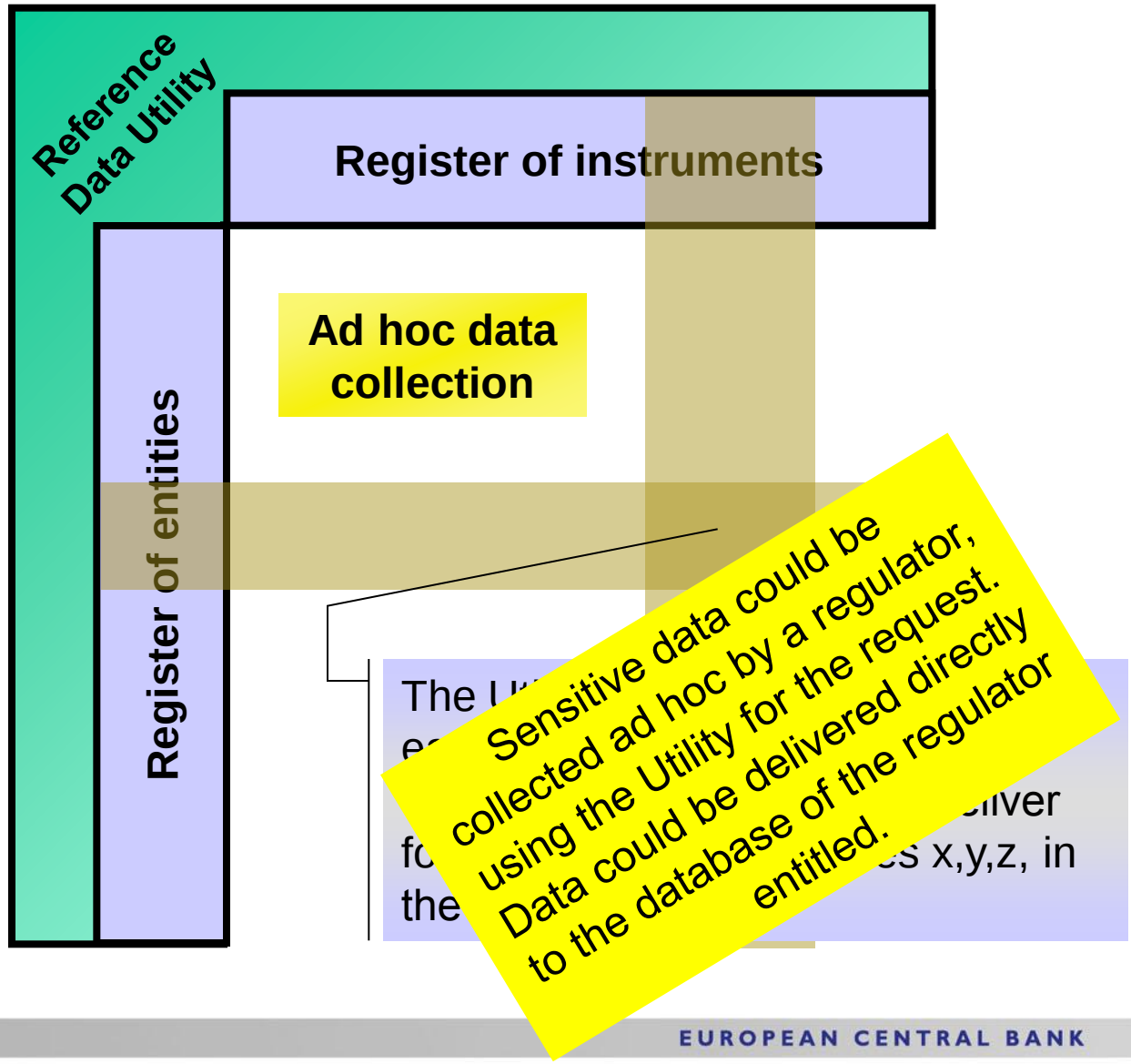
The Utility supports systemic risk analysis

Sampling becomes a low risk option for speed and cost control



The Utility supports systemic risk analysis

Confidential data could be handled in the same, safe way as today, but faster and more cheaply



The international aspect

Global Utility vs. National Law: an option

International Community

e.g. G20:
- Discusses new regulatory framework for financial markets
- Defines principles / goals for data

International Institutions

Governance of the Operational Entity
- Global „tour de table“ (IMF, BIS, industry, etc.)
- Establishment of Int’l Operational Entity
- Seed funding of Int’l Operational Entity ?

National Constituency

National Legislator

Issues law:
- mandates national authority
- empowers it to enforce the process and
- to farm out operations to an international entity.
- (EU issues specific EU law)

National Authority

Executes the legal mandate:
- Farms out operations to the Int’l Operational Entity
- Monitors compliance
- Enforces, applies sanctions

Entity

Complies with national law:
- Delivers and maintains data in the Utility as required, possibly using services.

International Operational Entity

Utility

Service agreements with national authorities

Runs the service:
- Collects data
- Distributes / sells data
- Certifies analysts
- Monitors compliance
- Informs national authorities
- Releases new standard items

Standards College

ISO

Develops/maintains standards:
- Designs initial standards
- Monitors market developments
- Steers evolution of standards
- Designs new standard items

International Utility: modular growth to global

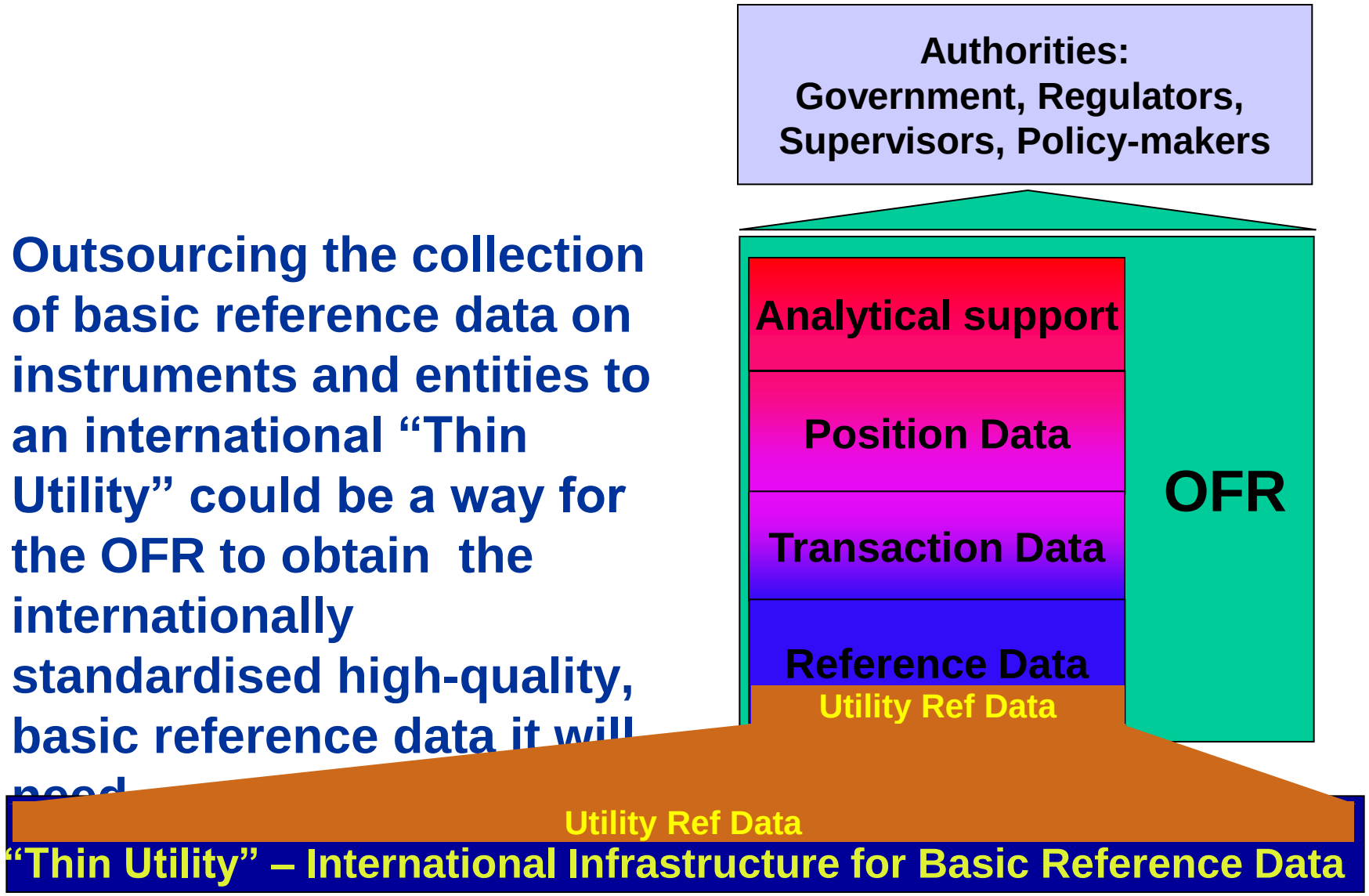
An international Utility could be established among a few leading financial markets.

It could grow from there in a modular fashion, further countries joining as they see fit, having passed adequate legislation.

International mechanisms and industry pressure could encourage further growth, once benefits become more broadly visible.

How could it fit in the case of the USA?

Outsourcing the collection of basic reference data on instruments and entities to an international "Thin Utility" could be a way for the OFR to obtain the internationally standardised high-quality, basic reference data it will need



Positioning and Design of a Reference Data Utility

Utility: a hub in the data ecosystem

A Utility offering standardised, complete and reliable basic reference data on instruments and entities could play the role of a bridge between

- **Registers, public and private**
- **Standards spaces**
- **Regulatory and commercial databases**
- **Collaborative services (an emerging category)**
- **Classical proprietary markets for data and information,**

It could play that role at an international level

A Utility could provide an anchor, a hub for connecting many segments of the international data ecosystem

Utility: an infrastructure for regulators

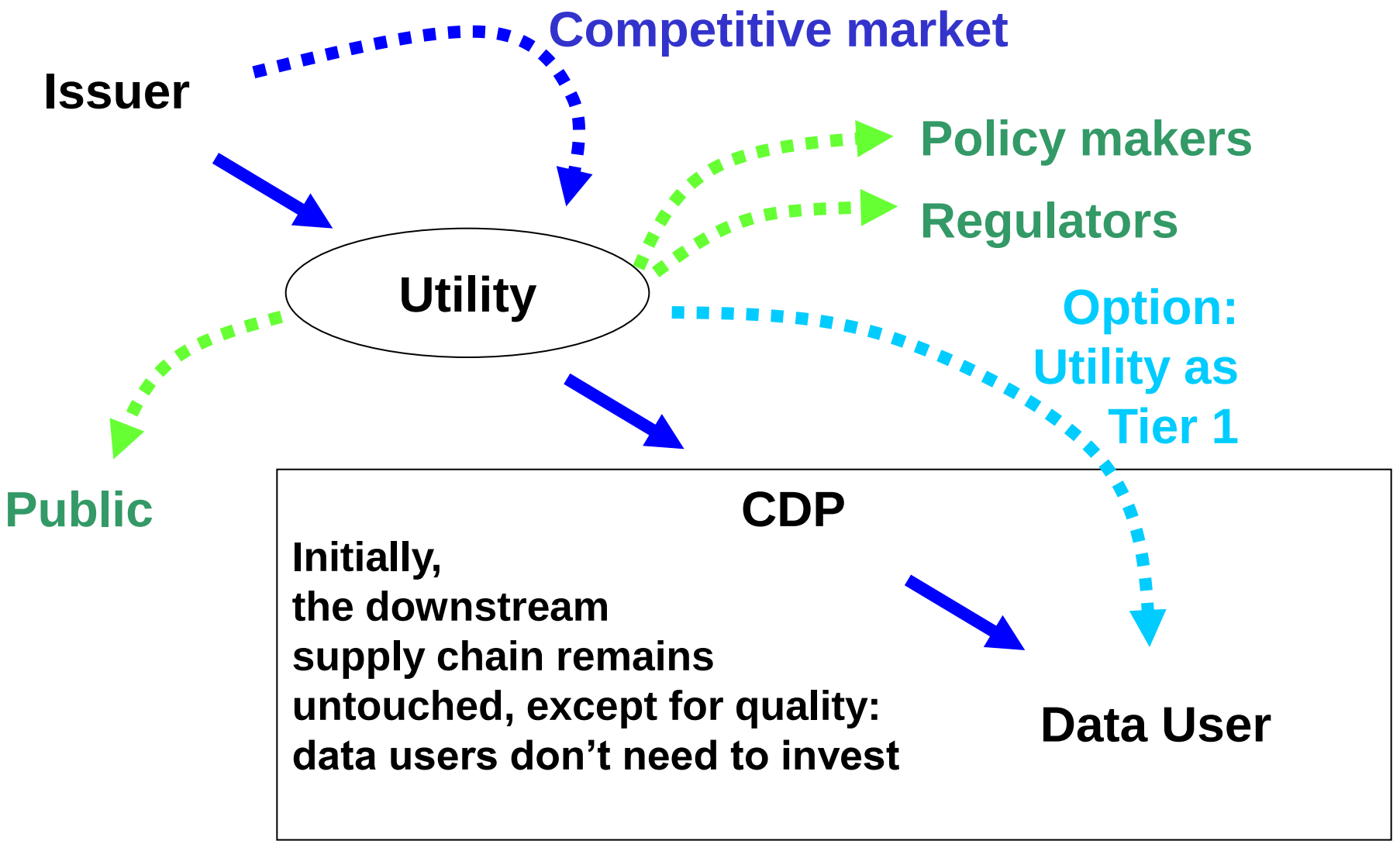
A Utility offering standardised, complete and reliable basic reference data on instruments and entities would offer an infrastructure enabling regulators to

- **Organise their own ongoing micro-data collection,**
- **Run fast, targeted ad hoc micro-data collections,**
- **Exchange micro-data with other regulators when needed**
- **Enable data fit for large-scale computing,**

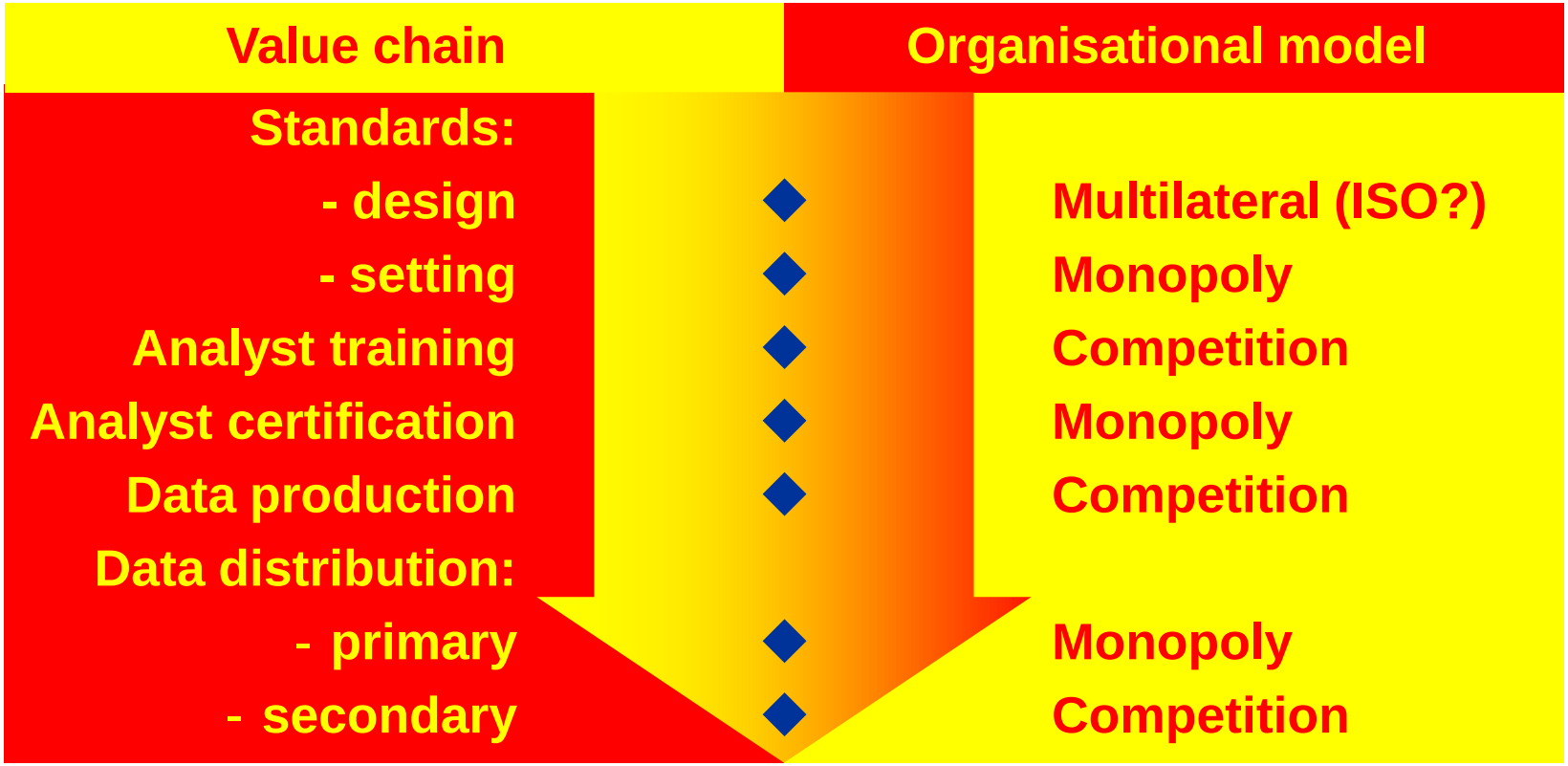
at an international level.

A Utility could be a shared, international infrastructure for regulators.

Positioning in the data supply chain



Utility value chain: monopoly vs. competition



Each stage of the value chain should be given the most suitable organisational model