

# Competency Framework

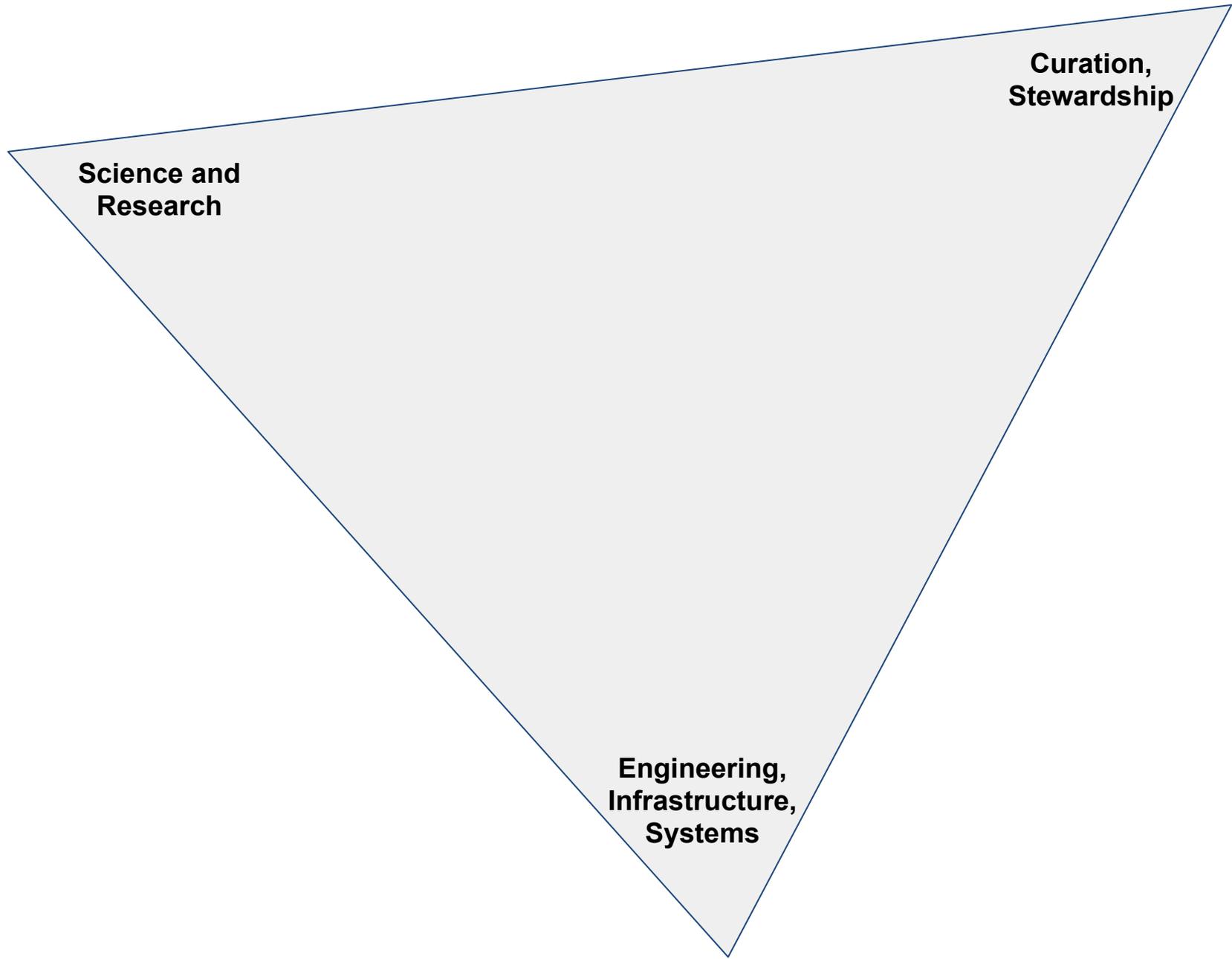
*Engineers, Statisticians, Data Scientists, Librarians,  
Data Curators & Researchers*

Wim Hugo  
CDIO: SAEON  
ICSU-WDS SC Vice Chair  
November, 2017

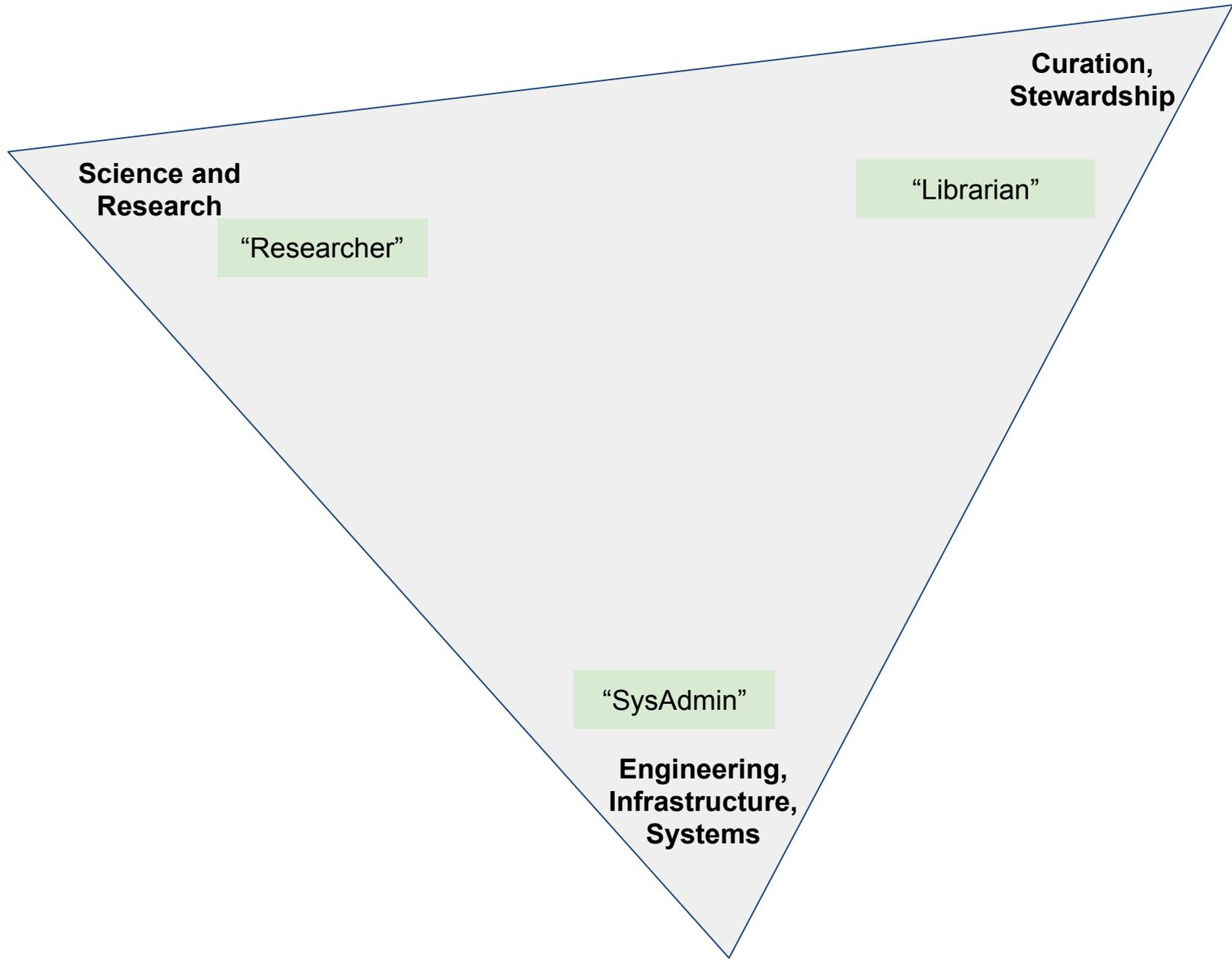


## Part 1: What Kinds of People Do We Need?

# Scope of Skills and Knowledge



# Scope of Skills and Knowledge



**Science and  
Research**

"Researcher"

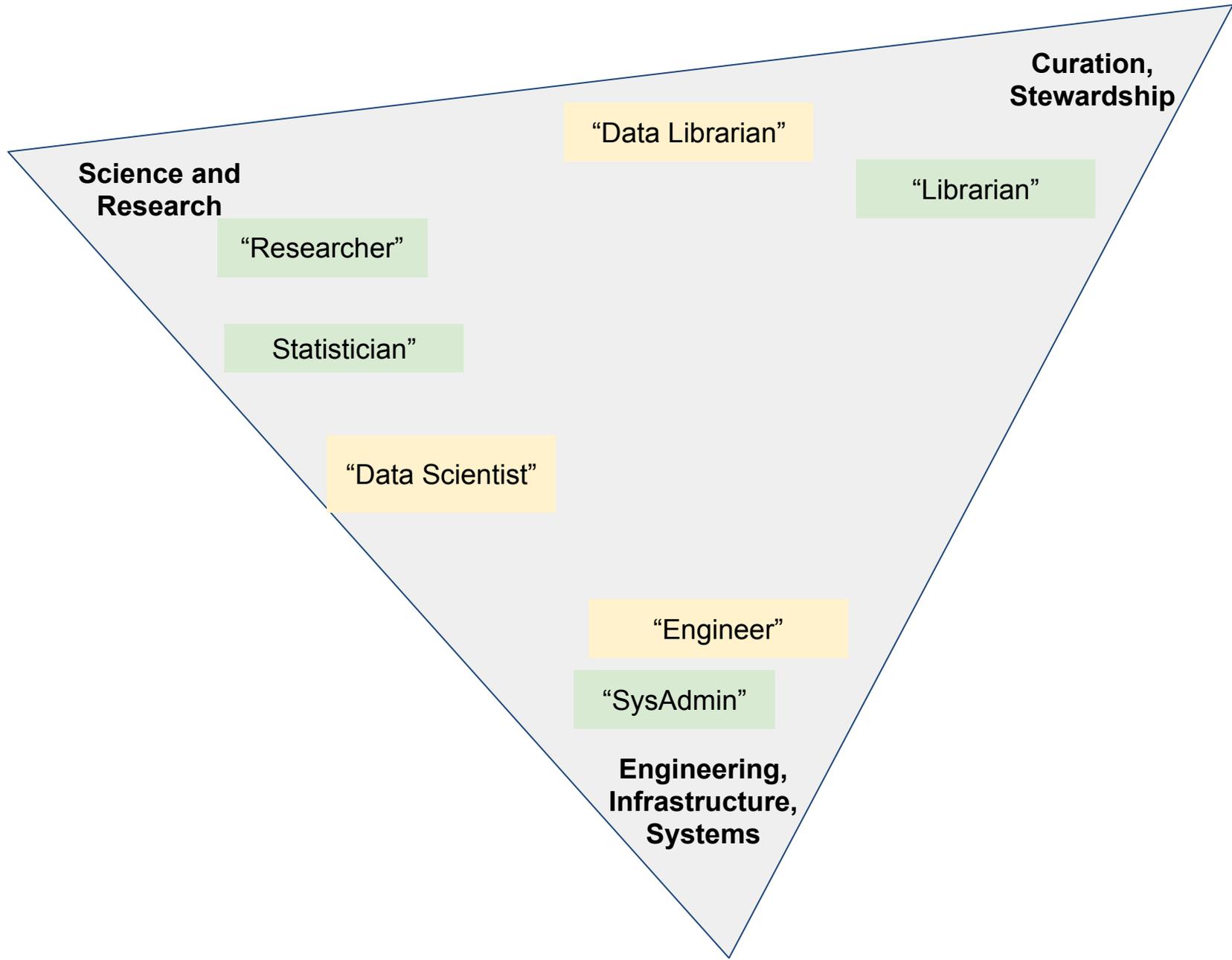
**Curation,  
Stewardship**

"Librarian"

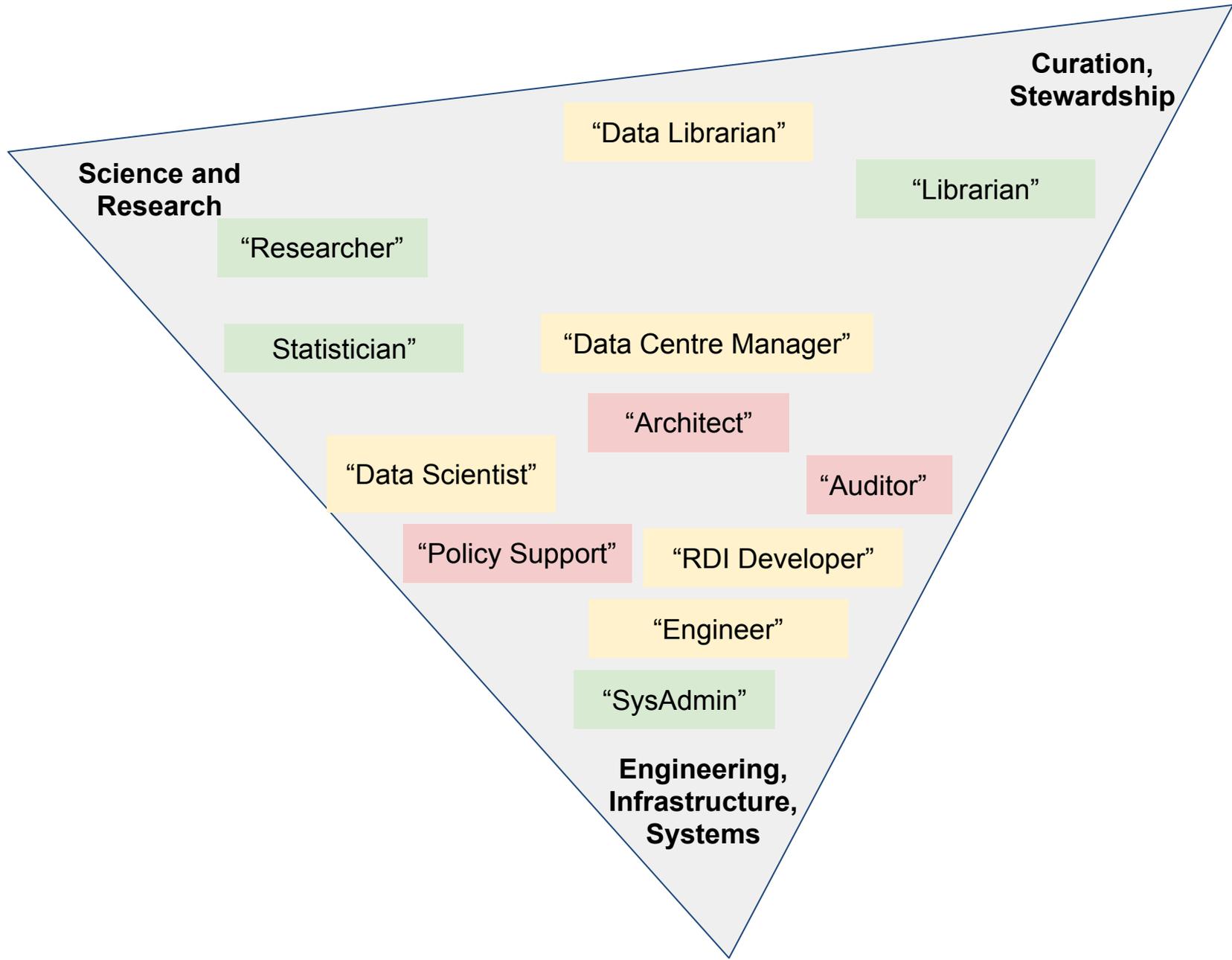
"SysAdmin"

**Engineering,  
Infrastructure,  
Systems**

# Scope of Skills and Knowledge



# Scope of Skills and Knowledge



Part 2: What do we need to know about?

# RDI Community Considerations

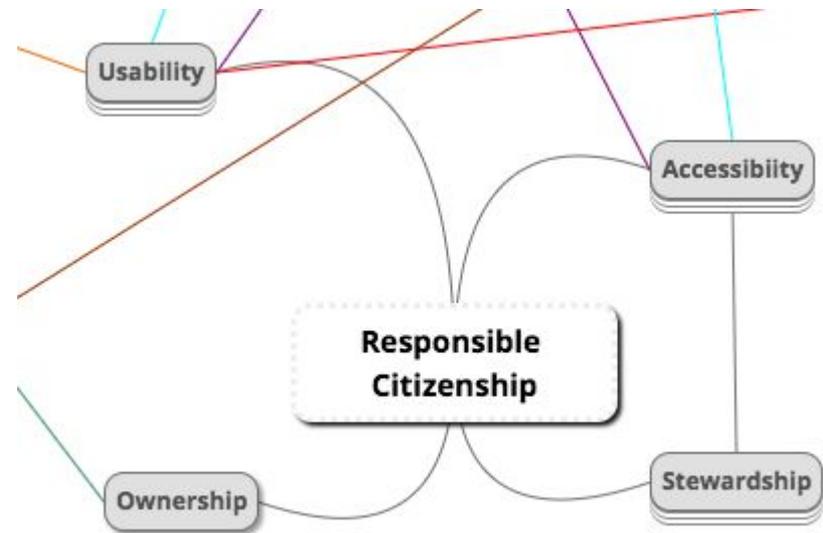
RDI is currently constructed within a framework of global initiatives and established community consensus. These includes, *inter alia*:

- The [FAIR](#) data principles;
- [GEO](#) Infrastructure, the [GEOSS Data Management Principles](#), and supporting [Implementation Guidelines](#);
- ICSU-World Data System [guidance on certification](#) as a Trusted Repository, developed in conjunction with [DSA](#) and governed by the recently constituted [CoreTrustSeal](#) Board;
- Published and emerging guidance within the broader [Research Data Alliance](#) collaboration, for example
  - [Dynamic data citation](#);
  - [WDS-RDA Working Group on Data Fitness for Use](#).
- Emerging practice in respect of [Essential \(Standard\) Variables](#).

The typical stakeholder groupings are **Individual Researchers**, their associated **Institutions**, Initiatives or Projects, **Repositories** and Data Centres, and **Community Initiatives** and **Global Authorities**. For each of these groupings, it is possible to define sets of guidance and best practice that are collectively referred to as the basis of **Responsible Citizenship of the Data World**.

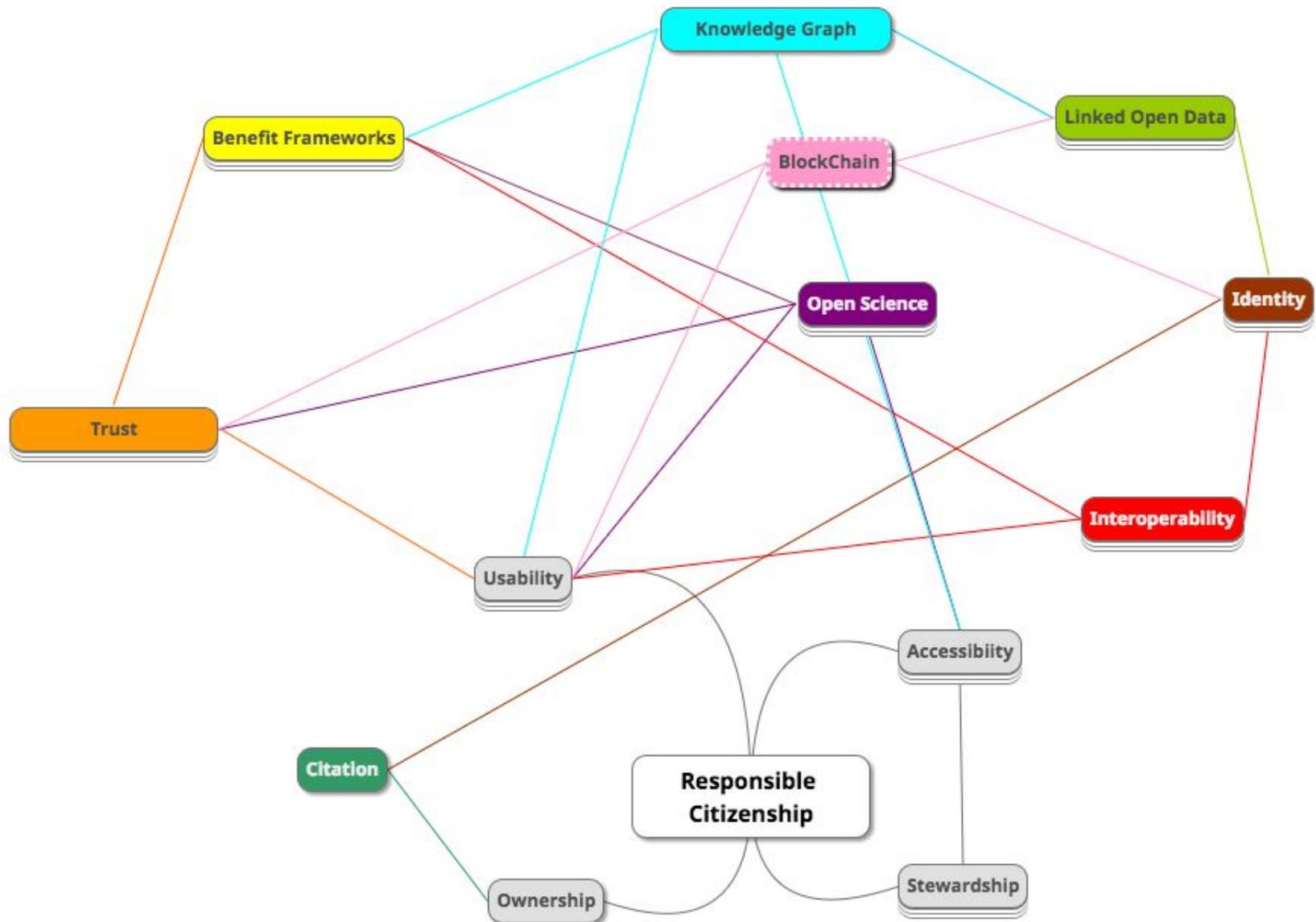
# Separation of Concerns

The main focus of efforts such as FAIR and GEO is to define the **'Performance Parameters'** under which data (and by extension any other research output) will be optimally produced and re-used.



- Parameters: Usability, Accessibility, Ownership, Stewardship
- Superimposed on these are **'Concerns'**

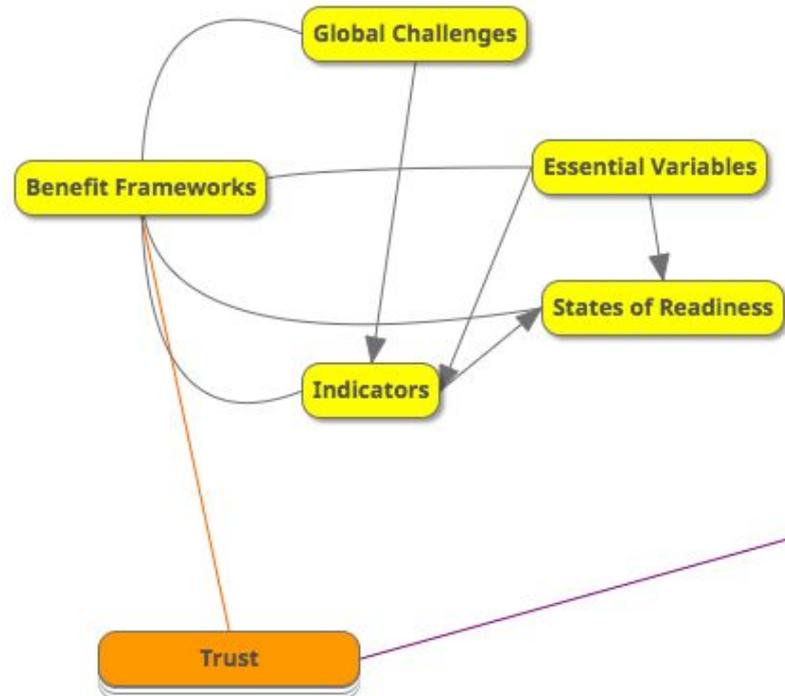
# Scope of Design Considerations for RDI





# Benefit Frameworks

Development of appropriate portfolios of Essential or Standard Variables that allow integration of data across temporal and spatial scales, across biomes, and across regions.

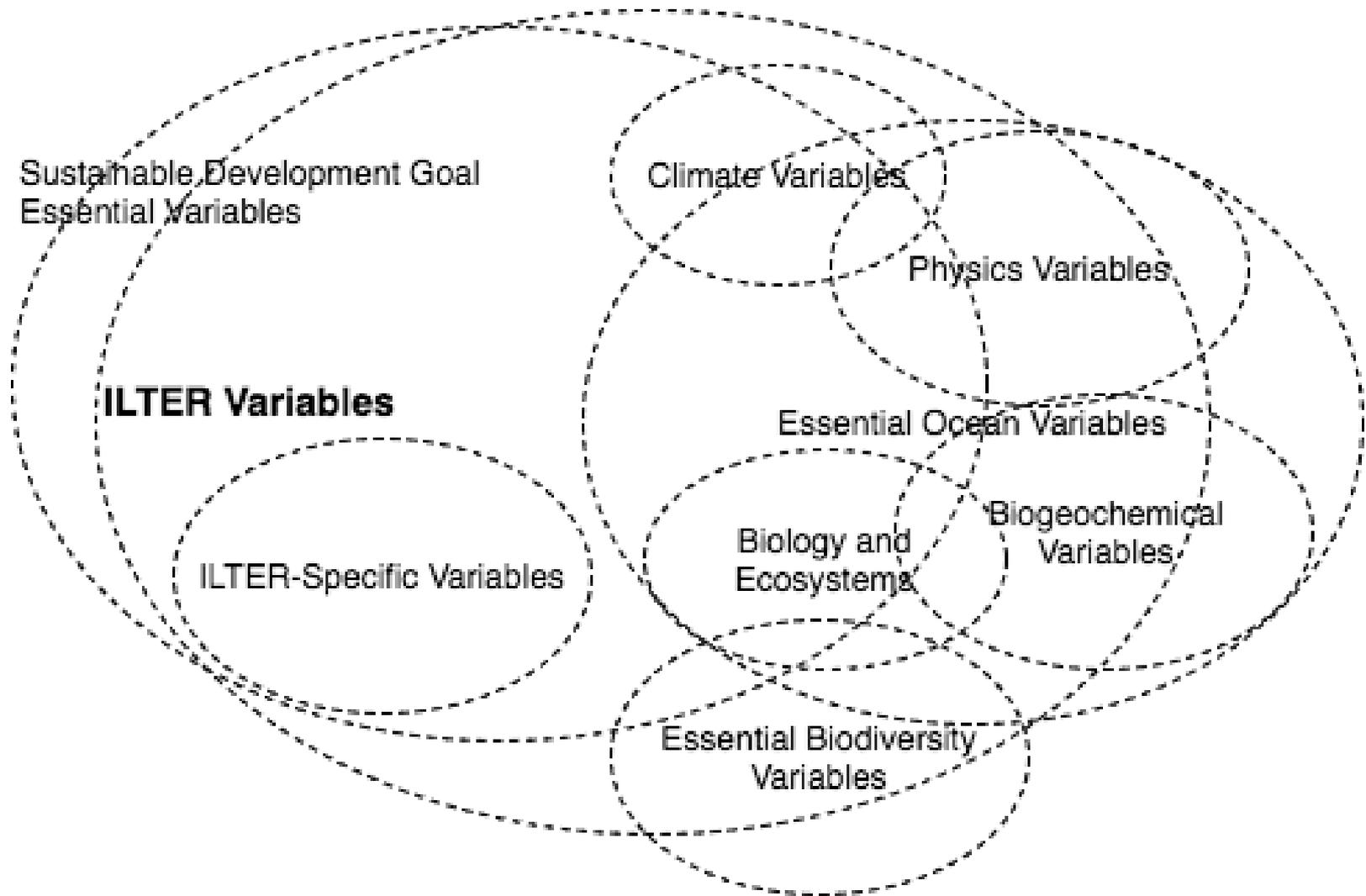


- Existing and emerging portfolios to be taken into account (EBVs, Ocean Variables, Climate Variables, Fluxes, ...)
- Elements of scientific relevance, protocol, semantics and vocabulary
- Define appropriate 'Data Families' and associated interoperability standards
- Consider mapping to global indicator portfolios (SDG, Sendai, Aichi, ...)
- Formalise 'State of Readiness'
- Critical for transdisciplinary research

# Data Service 'Readiness' - Value Chain

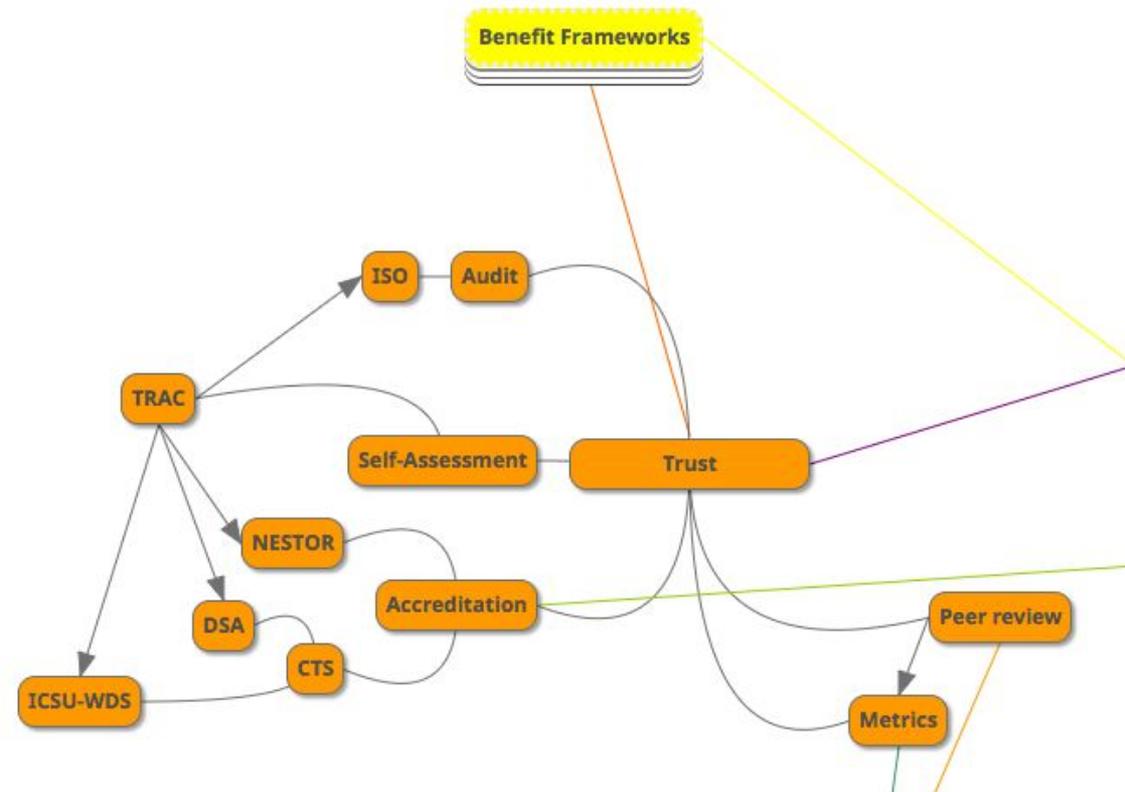
- “Raw Data” - data for which no quality assurance, correction, or other form of preparation has been done, but discovery metadata may be available;
- “Publication-ready Data” - for which appropriate **quality assurance** has been done, and citation, administrative metadata is additionally available;
- “Analysis-ready Data” - for which usability metadata is available, and community conventions in respect of vocabulary (semantic), and structural (schematic) interoperability conventions have been applied. Typically, portfolios of **Essential Variables** fall into this category.
- “Indicator-ready Data” - for which in addition to the above, syntactic (service) interoperability is also available and data can be included in **trans-disciplinary indicator** combinations. In such cases, it is very useful if semantic annotation is applied to contextualise the application of one or more Analysis-Ready data sources in the derivation of an Indicator.

# Essential Variable Landscape



# Trust

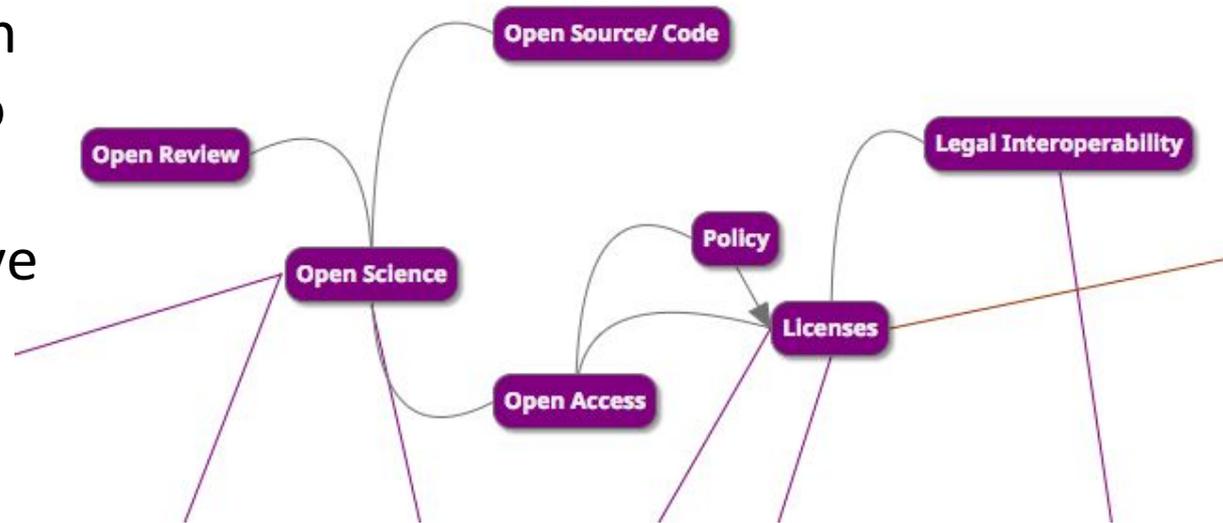
Trust is at the centre of the process of science, and increasingly important in grant-funded research. Extensions beyond data to publications, software, vocabularies, ...



- Network-level membership of ICSU World Data System
  - Accreditation of network members
  - Individual certification of data centres and repositories using CoreTrustSeal
- Dataset-level quality assurance and metrics in collaboration with GEO
- Potential future implications of BlockChain technology for trust and provenance
- Critical for automation

# Open Science

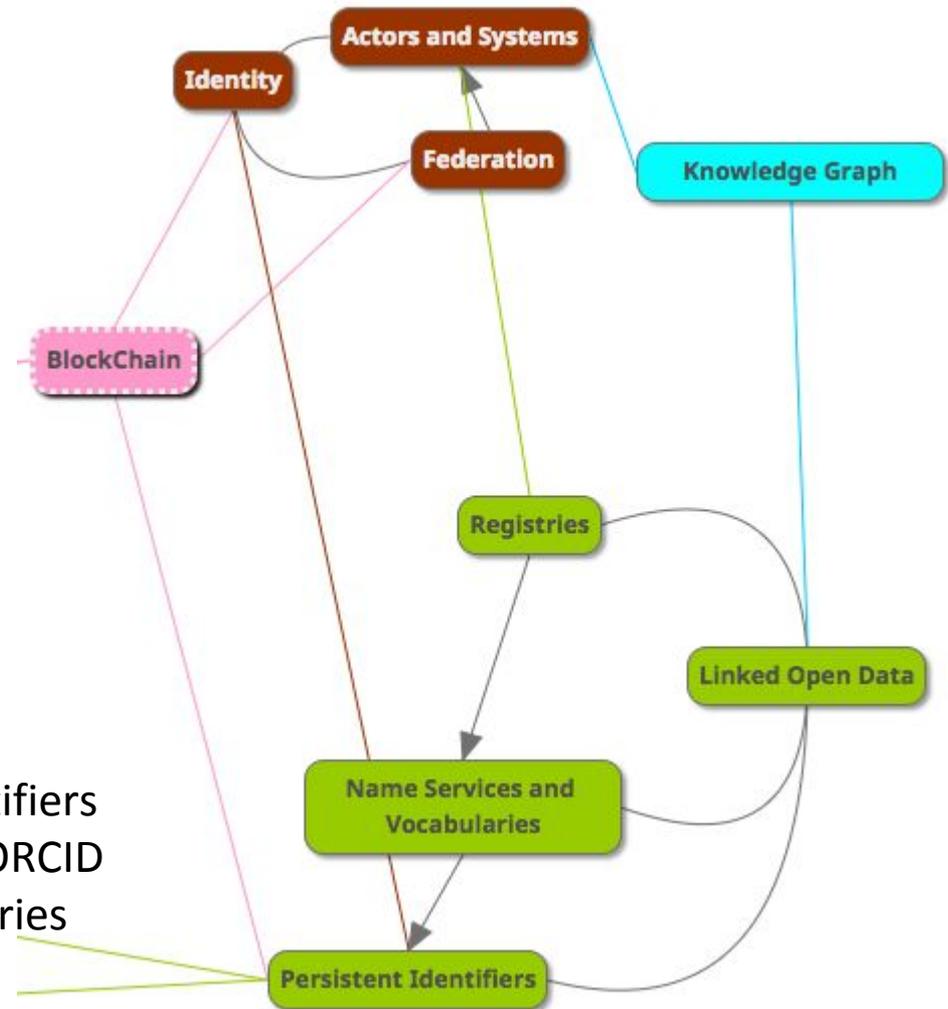
Open Science is driven largely by a desire to make science more useful and to improve **validity**



- From RDI perspective, policy is not important, licenses are
- Embargo periods and common sense
- Creative Commons Licenses preferable
  - Machine-readable exceptions to CC Licenses
- Legal Interoperability Framework needed
- Again, critical for automation

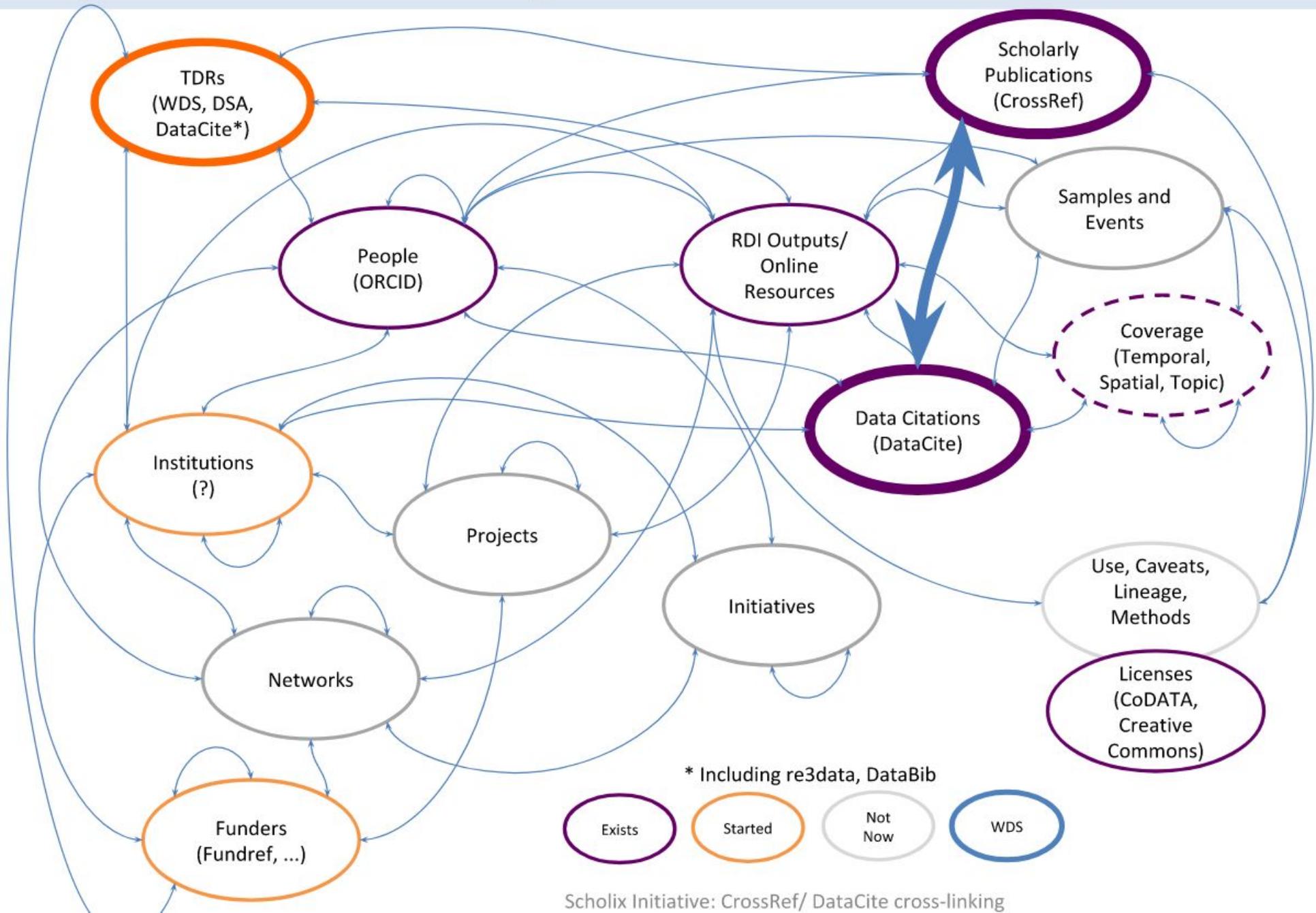
# Identity and Persistent Identifiers

Identity is increasingly important for the actors and systems in Research Data Infrastructure, and underpins necessary constraints on the Semantic Web



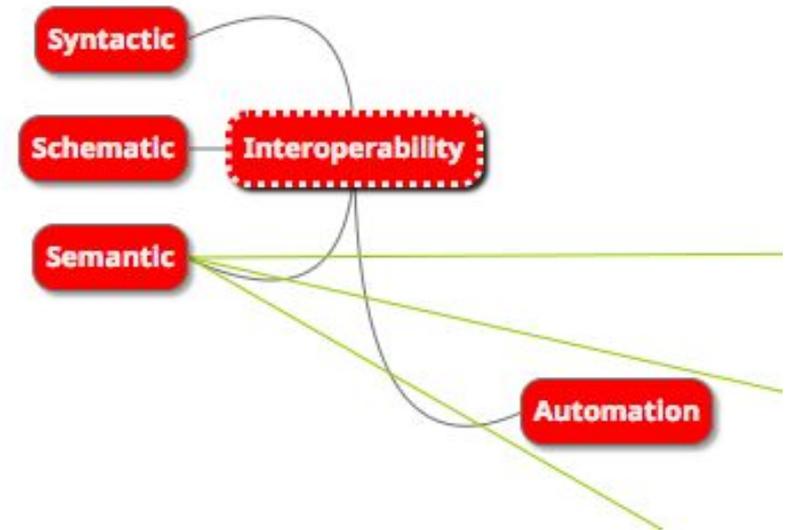
- Promotion and use of Persistent Identifiers
- Federated, single sign-on - EduGain/ ORCID
- Agreement and guidance on vocabularies
- Registries of important elements
  - Especially Sites
- Citation of dynamic data sets
- Potential Blockchain implications for Identity and Federation

# ICSU-WDS Knowledge Network: the Fabric of Science



# Interoperability

Interoperability is directly linked to readiness of data and variables, automation and machine-readability, and application outside discipline-specific silos.



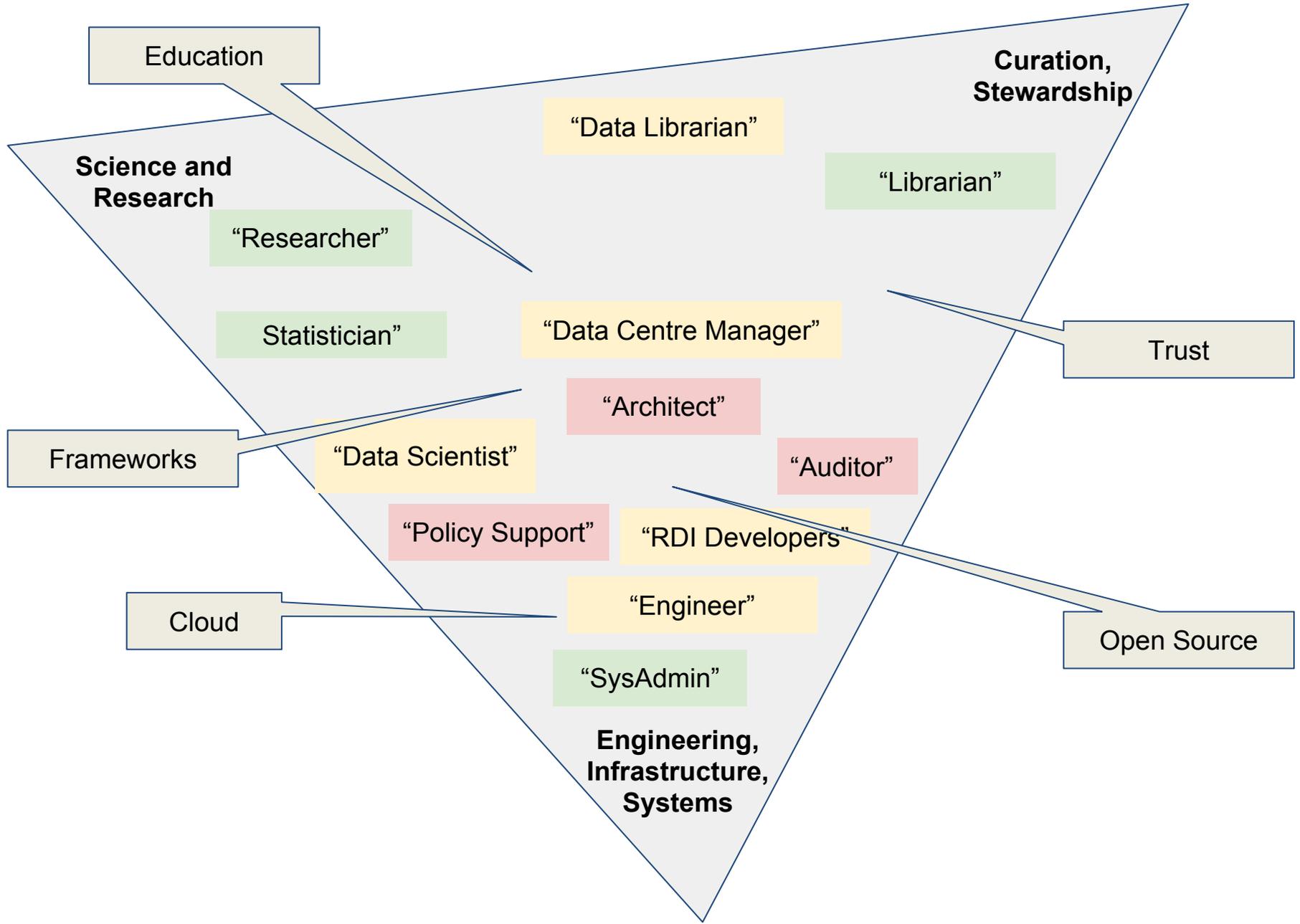
- Development of essential variables for semantic interoperability
- Agreement on metadata and data schema, services based 'Data Family' approach
- Brokering Framework - extending [Scholix](#)
- Loosely coupled architecture required ...
- Automation must always be in mind

## Part 3: What do we need to focus on?

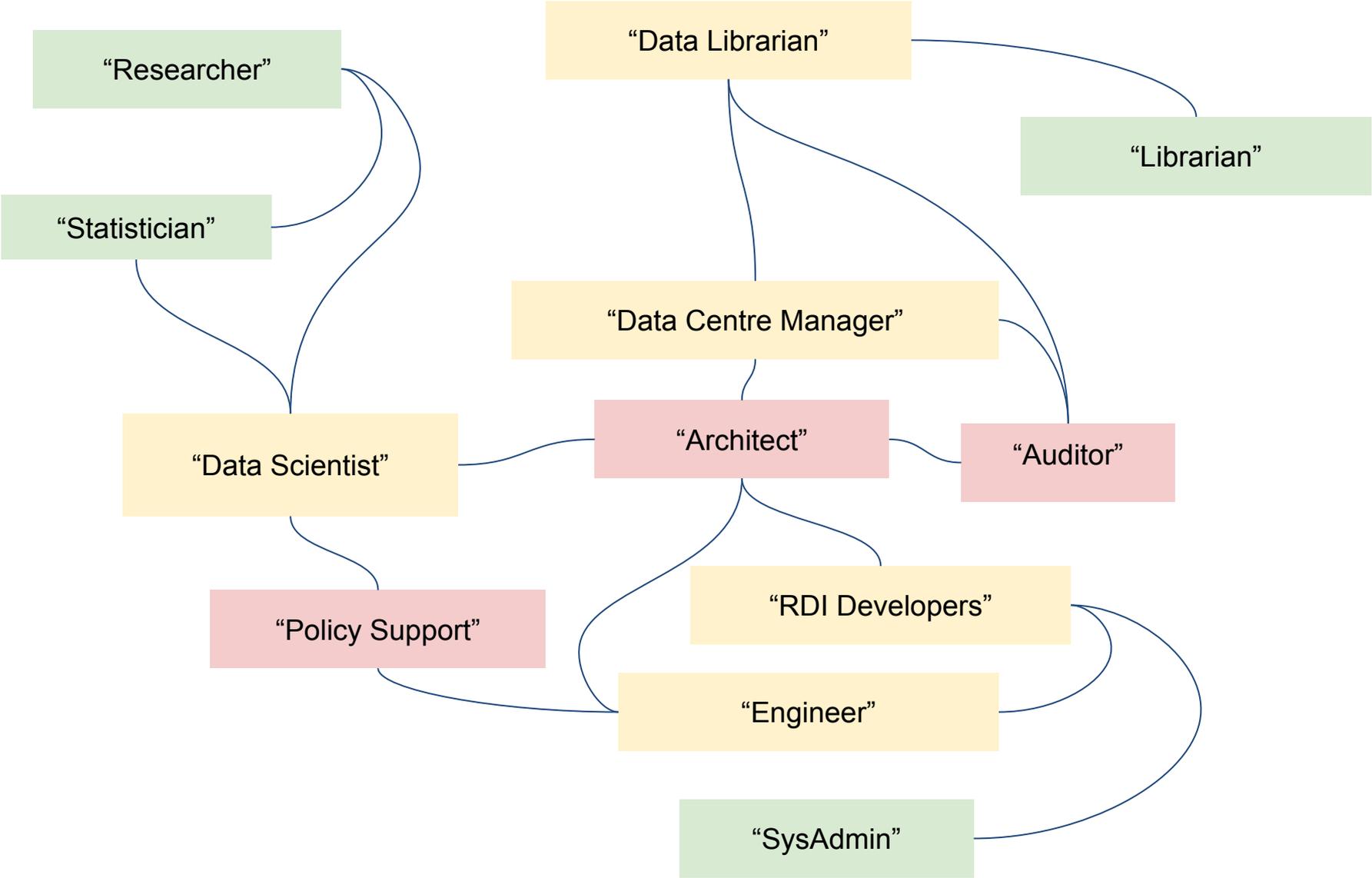
# Human Resource Challenges

- Deficit in Multi-disciplinary Skills
  - Auditors
  - Data Librarians
  - Policy Support
  - Data Scientists
- Data Centre Managers and Architects
  - Can supplement skills
- Critical Mass and Promotion Prospects
- Measurement and Compensation
  - Science is measured (largely) via publication
  - Engineering is measured by risk management
  - Curation is measured by permanence

# Solutions



# Solutions: Career Pathways



# Responsible Citizenship

Metadata is the tax of the data world.  
It is not pleasant to pay it, but  
citizens cannot expect benefits (infrastructure, services) if  
they do not pay their taxes ...

