

# Challenges for Multi-Disciplinary Research Data Infrastructures

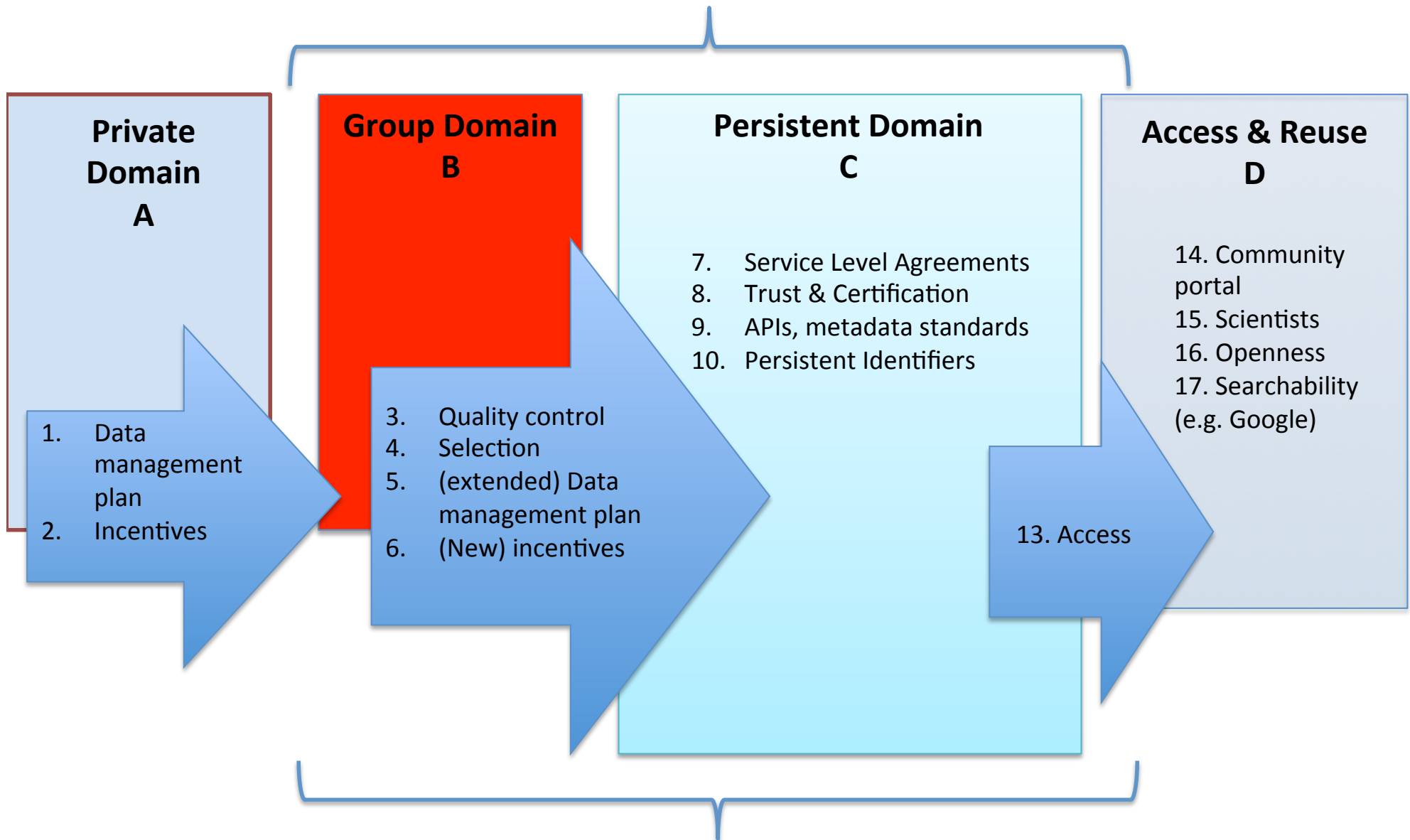
## The Group Domain

Harry Enke, Jochen Klar  
Leibniz Institut für Astrophysik Potsdam  
Radieschen  
[forschungsdaten.org](http://forschungsdaten.org)



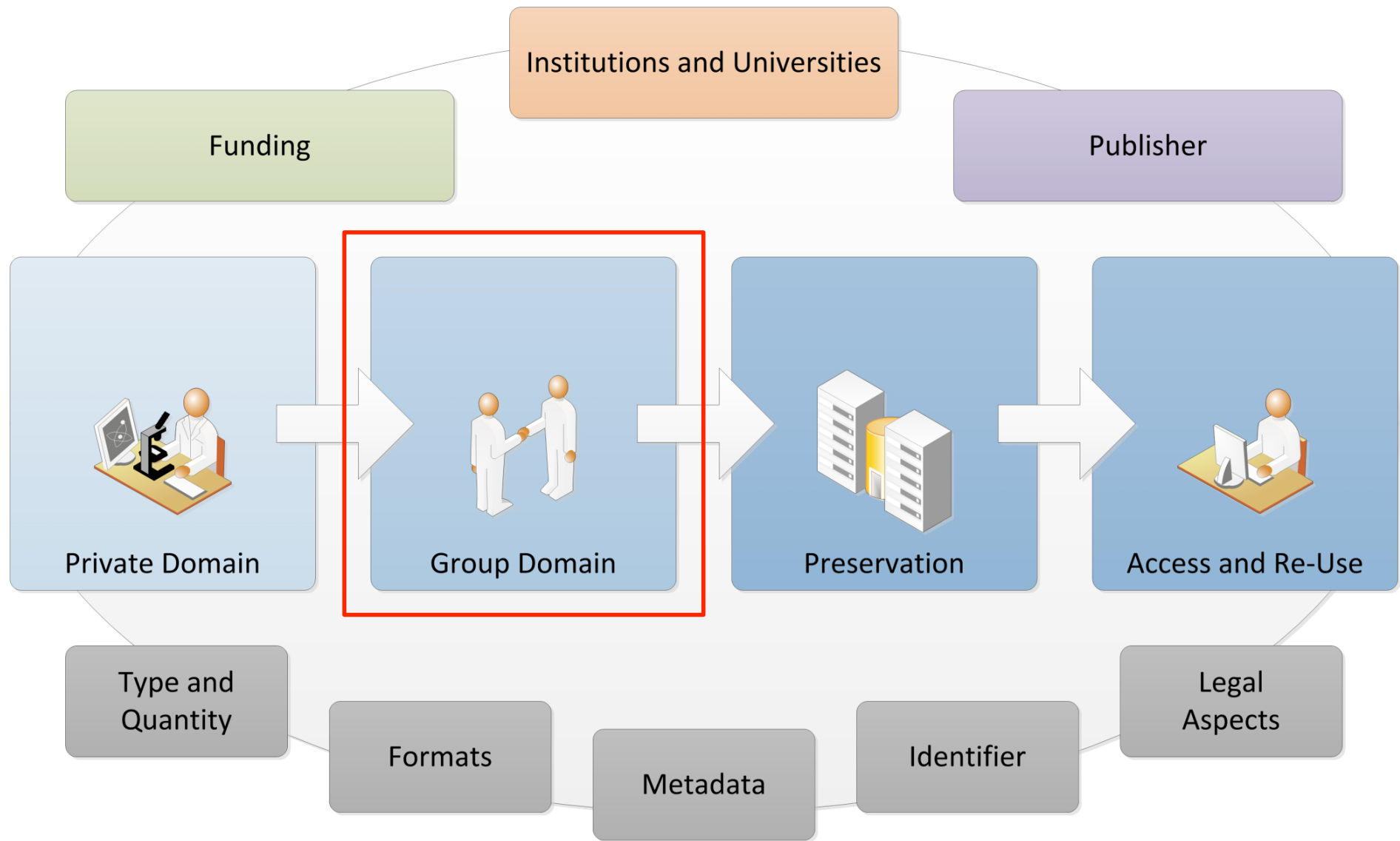
# Domains I

11. Preservation Policy



12. Community-specific organisational structures

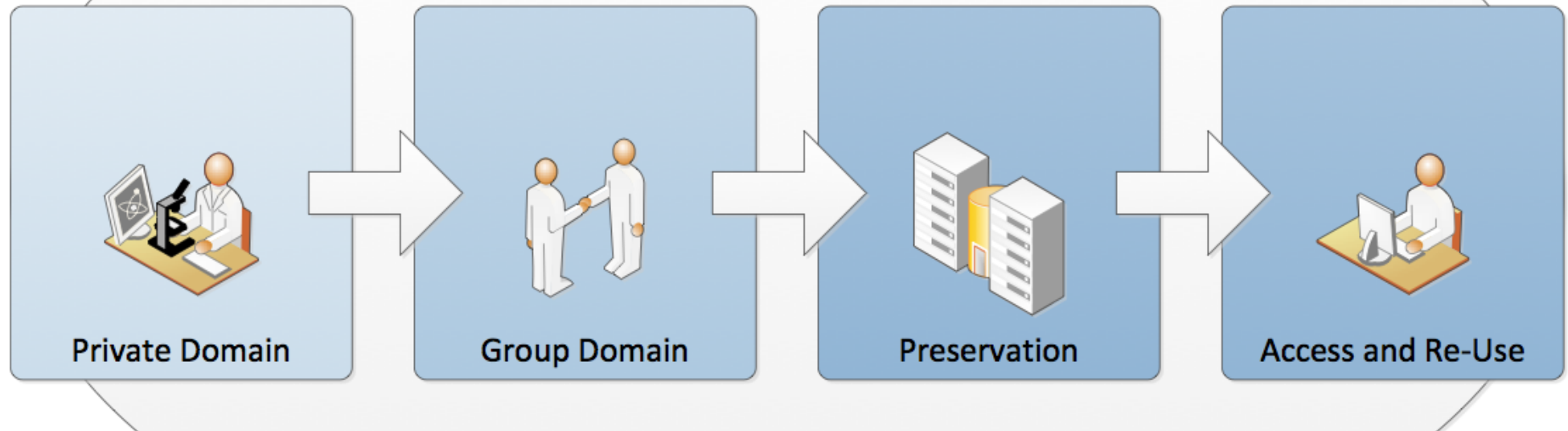
# Domains II



Funding structure, organisation of research is country specific  
Obstacles for VRE

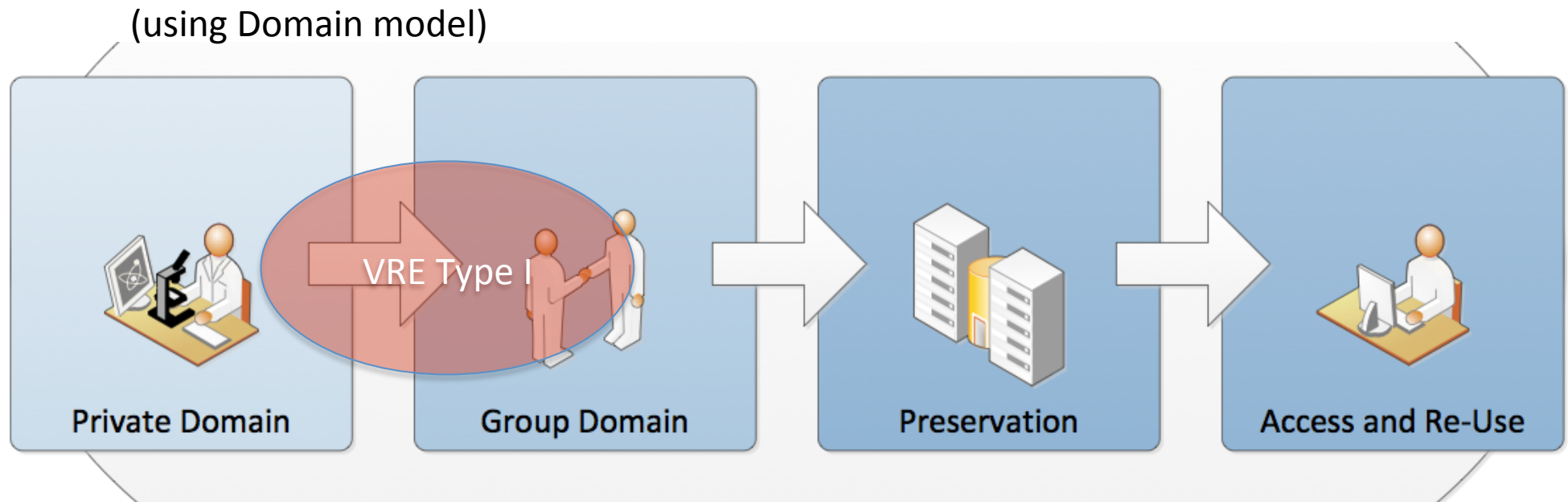
# Locating Workflows in Research Process

(using Domain model)



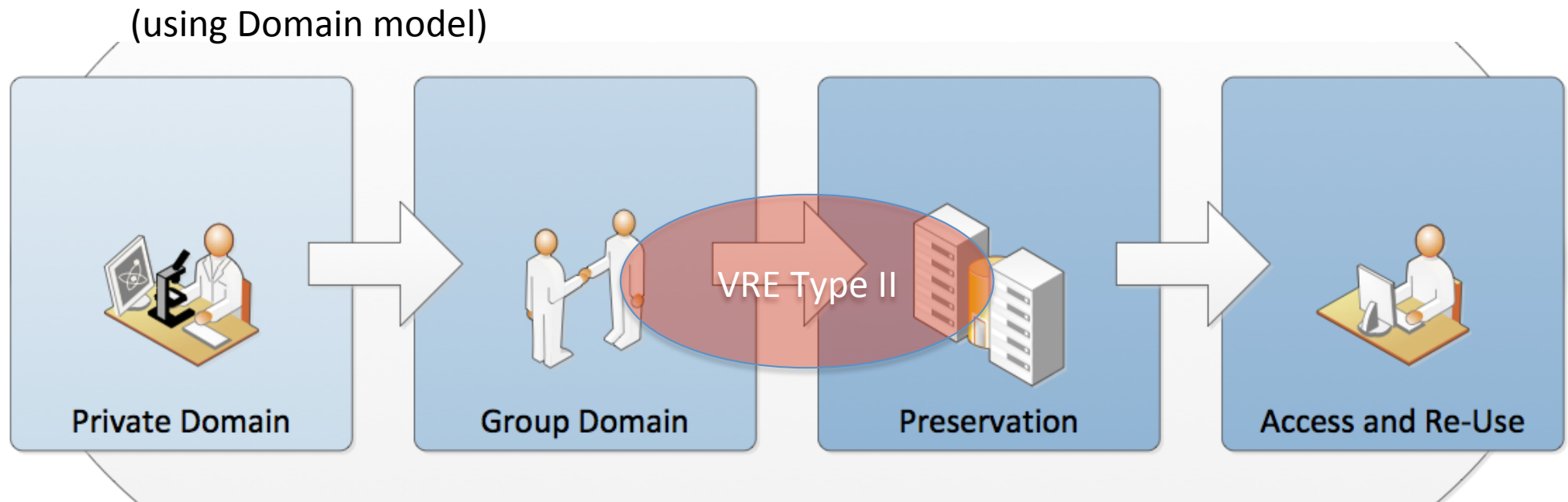
- **Private Domain**
  - Generation / Processing of research data not separated
- **Group Domain**
  - Collaboration working on common set of data, workflow is separated into specialities
- **Preservation Domain**
  - Archival of data, workflows further differentiated, include curation procedures still oriented on discipline specifics
- **Access & Re-Use Domain**
  - Publishing for broad audience, more curation, separation from producing scientists, no scientific processes on data

# Locating VRE in Research Process



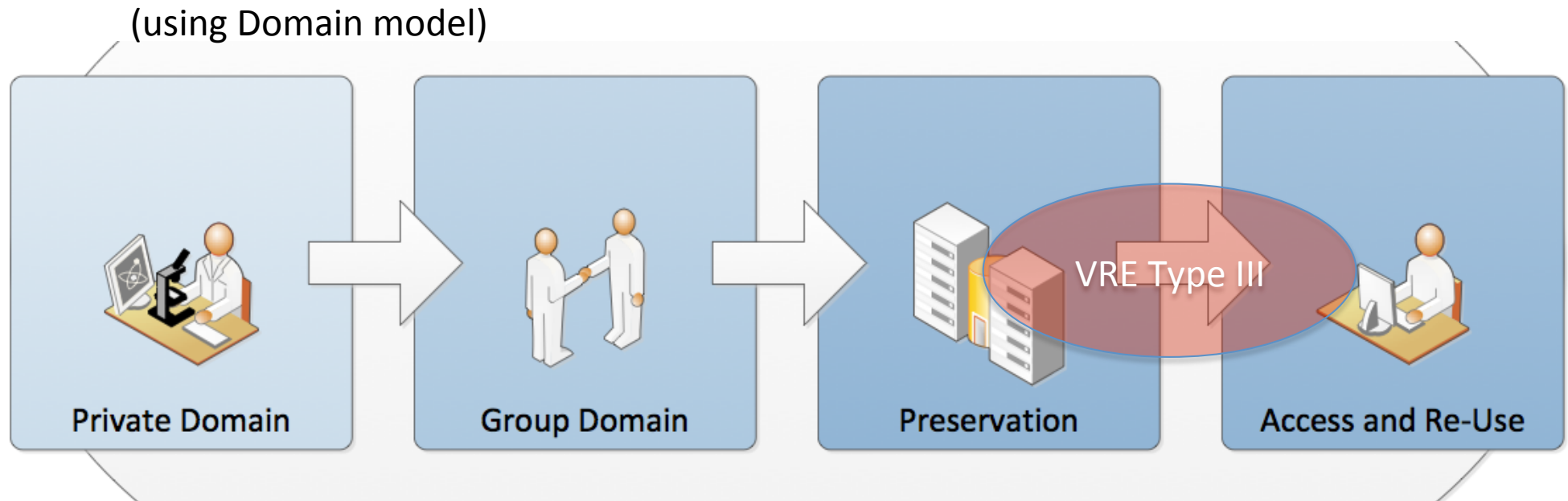
- **VRE Type I (data controlled by science collab.)**
  - Managing data flow / work flow from Private to Group Domain:
  - Required:
    - Collaboration Space [common storage + compute resources, communications (wiki, mailing list),.... ]
    - Discipline specific metadata sets
    - Tools for easy enrichment of metadata {,apps'}

# Locating VRE in Research Process



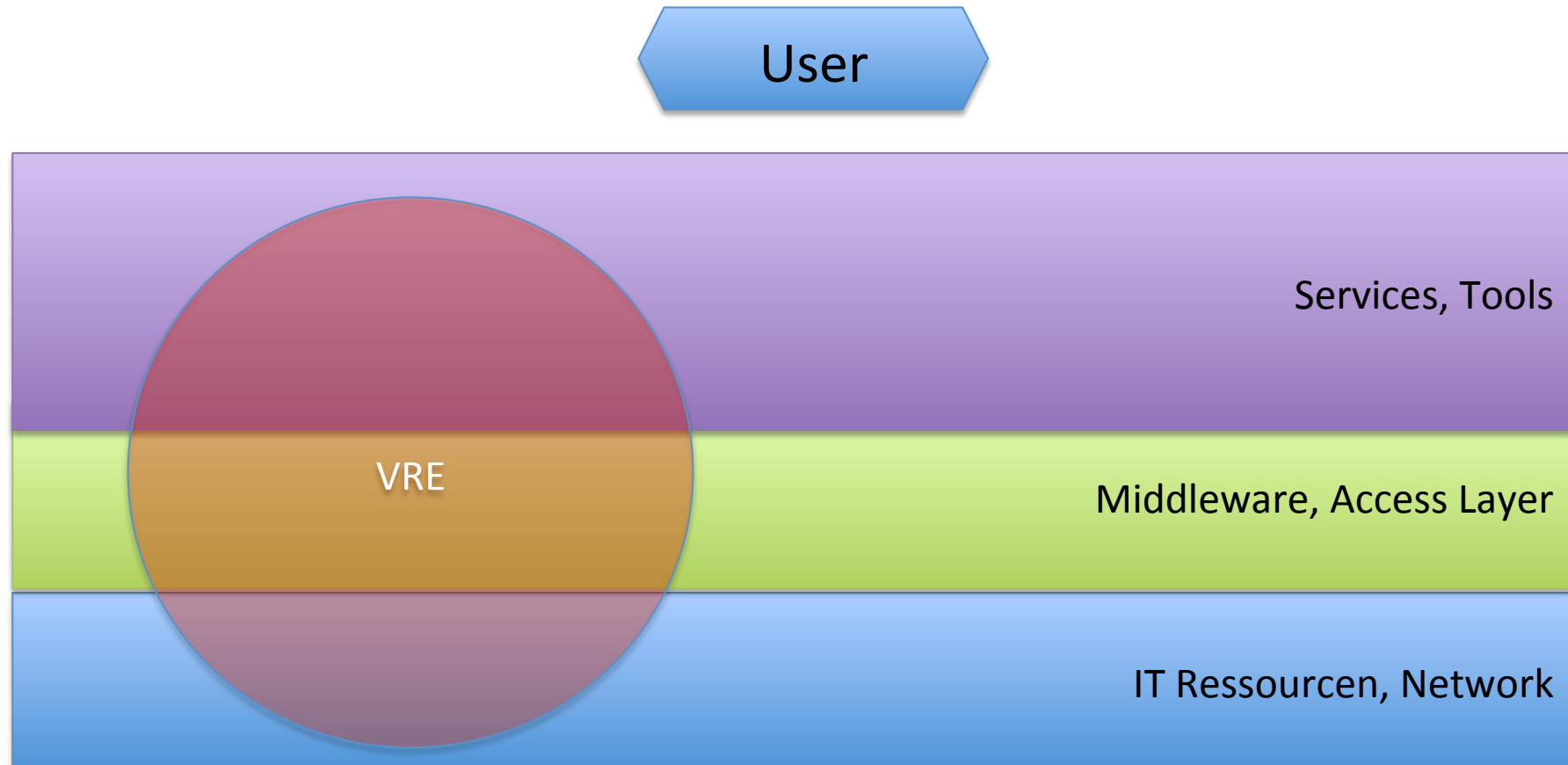
- **VRE Type II (data controlled still by science collab.)**
  - Managing data flow / work flow from Group to Preservation Domain:
  - Required:
    - Archival procedures: metadata, ingest & quality control, longterm storage
    - Webservices & tools for accessing, retrieving, working & enhancing data
    - Standards for webservices & tools
    - Collaboration Space [common storage + compute resources, communications (wiki, mailing list),.... ]
    - Discipline specific metadata sets
    - Tools for easy enrichment of metadata {,apps'}

# Locating VRE in Research Process



- **VRE Type III (data controlled by institutions)**
  - Managing data flow / work flow from Preservation to Acces Domain:
  - Required:
    - Archival procedures: metadata, ingest & quality control, longterm storage
    - Webservices & tools for accessing, retrieving working, enhancing data
    - Standards for webservices & tools
    - Collaboration Space [common storage + compute resources, communications (wiki, mailing list),.... ]
    - Discipline specific metadata sets
    - Tools for easy enrichment of metadata {,apps'}

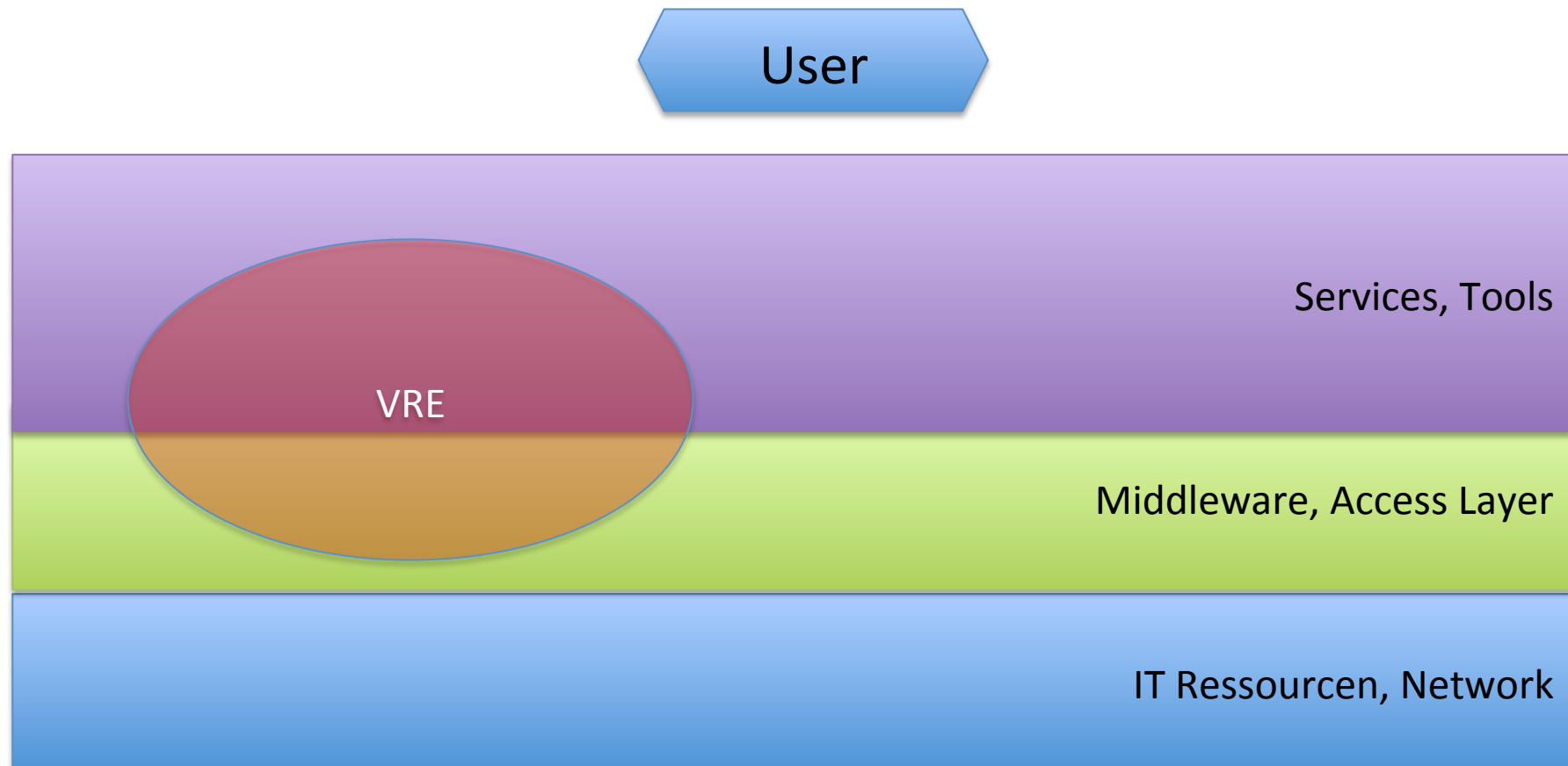
# Locating VRE in IT Infrastructure Layers



- VRE managing all layers : {example LHC grid, LOFAR}
  - Includes many different workflows, not confined to scientific workflows
  - Needs many additional management procedures
  - Often monolithic organisational structure
  - Huge data sizes, simple data structures

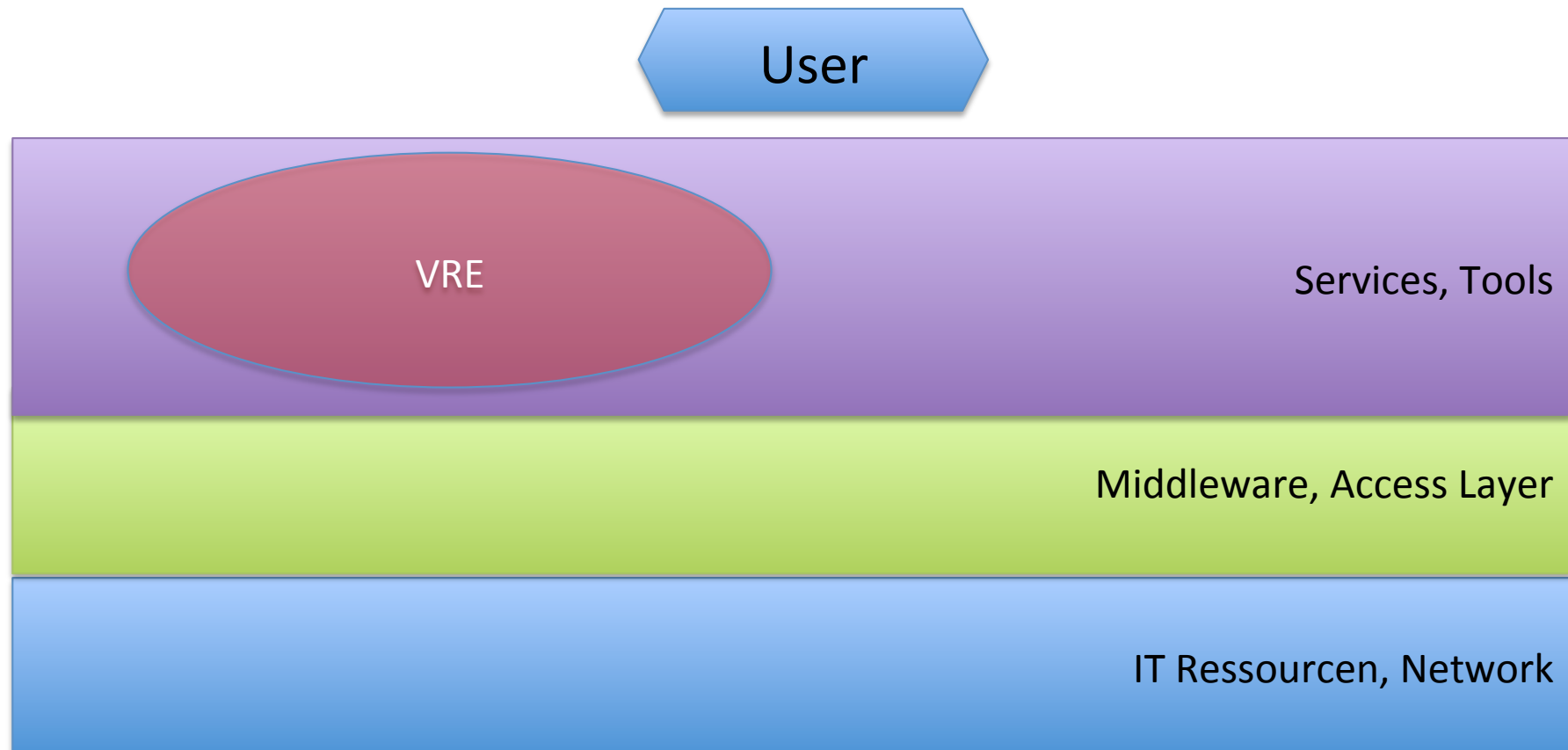


# Locating VRE in IT Infrastructure Layers



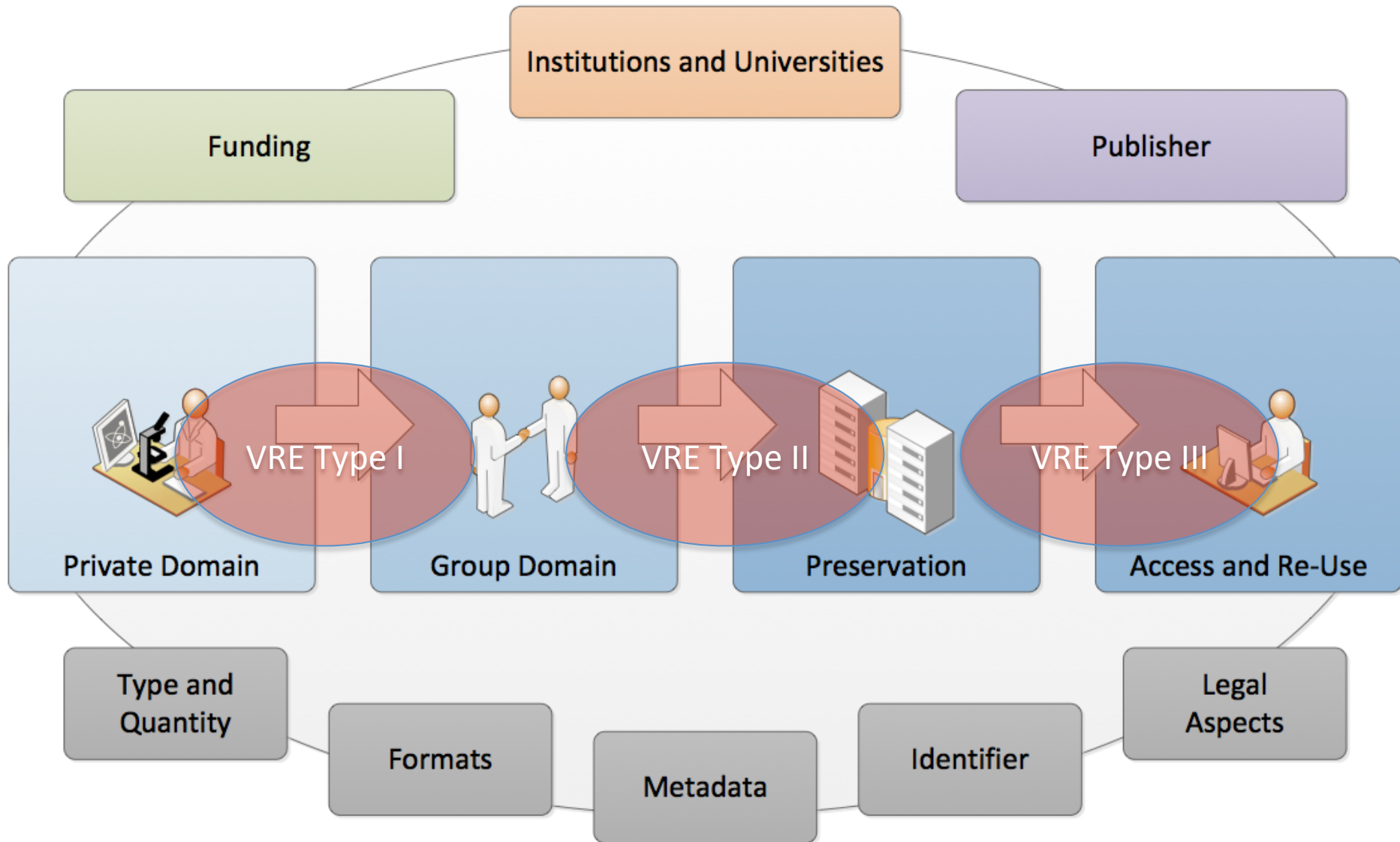
- VRE combining middleware & service layers : {example C3Grid, TextGrid }
  - Includes different workflows
  - Focus on few scientific tasks possible
  - Huge data sizes, more complex data structures

# Locating VRE in IT Infrastructure Layers



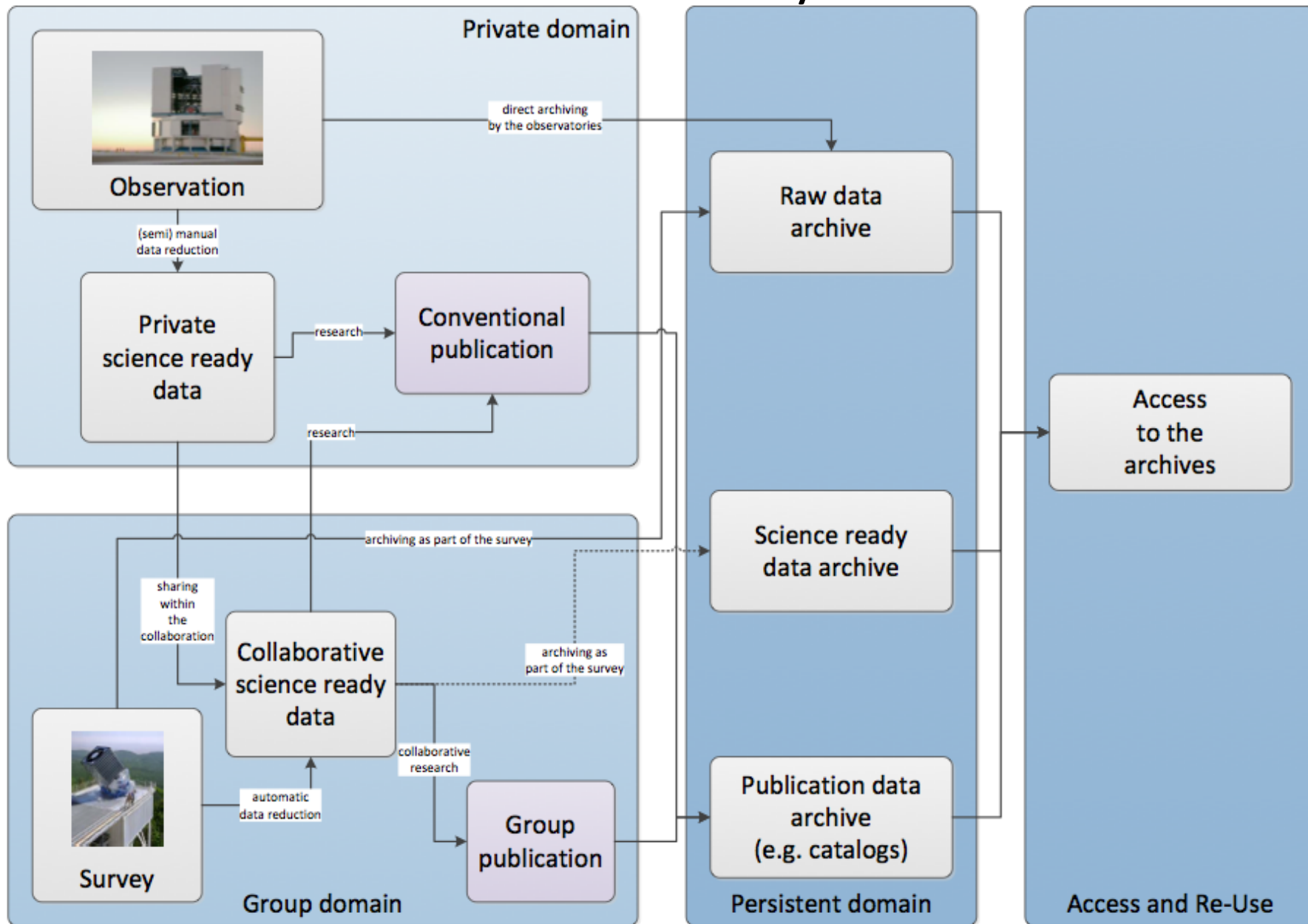
- VRE focussing on services & tools: {example: da | ra, data cite}
  - Focus on single or narrow set of tasks
  - Focus on few workflows
  - More complex data structures
  - Often meta task {improving useability of data}

# Workflow / Data Flow Analysis



VRE trying to cover all domains are bound to fail,  
Analysis of workflows is crucial

# Workflow / Data Flow Analysis Astronomy



# Workflow / Data Flow Analysis Social Sciences

Workflow for Statistical Analysis  
of data sets,  
core element for the VRE,  
valid for Private & Group Domain

Data stands for:  
raw data, re-used data,  
scientific use files, test sets,  
And syntax files for SPSS, Stata, R, ...

taken from:  
„Expertise zur Errichtung einer  
Virtuellen Arbeitsumgebung fuer die  
Sozial-Oekonomische Berichterstattung (SOEB)“

