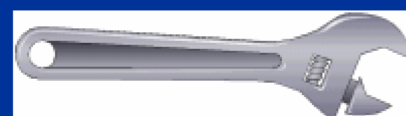
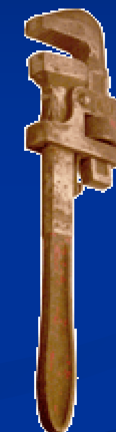
Software and Data Systems (SDS) Strategic Planning





Duane Edgington

Andrew Chase, Dan Davis, Kevin Gomes,
John Graybeal, Dorothy Oliver, Karen Salamy

Planning also included:

Jim Bellingham, Dave Caress, Mike Godin

“Need Data Monkey Wrenches” – *Steve Haddock*

“You don’t need a weatherman to know which way the wind blows” – *Bob Dylan*

What Is Our Status?

- What we have learned in our discussions with science (John)
- Planning as it stands and results of SWOT analysis (Dan)
- Potential directions that have emerged for the SDS group (Kevin)
- Issues to resolve and what we need to be successful in these directions (Kevin)

Science Discussions Background

- Spoke with most scientists at MBARI
  - Johnson, Clague, Vrijenhoek, Haddock, Paull (Ussler), Robison, Scholin, Chavez, Barry (Caress)
  - Couldn't get to everyone; some too busy
  - Discussed science themes and goals
- Scientists identified their/MBARI needs
  - Often claimed as community needs also
- Discussed all types of projects
  - Science Pull—Technology Push—Infrastructure
  - Note: Even scientists' needs may not be "science pull"
  - Engineering Goal: Move Technology Push and Infrastructure projects to the Science Pull category

What have we learned? Science input about…

- Strengths | Opportunities | Themes
- Community* Needs and Opportunities
  - Data Integration and Visualization
  - Quality Control
  - Data Organization (Management)
  - Tools & Applications
  - Event Response and Real Time Elements
- Our Approach

* Community = MBARI Science Community

What have we learned?

- Themes
  - Many science questions were multi-disciplinary
    - Example: Understanding material flux involves 3 types of transport, biological and physical contributions
    - Example: Understanding biological distributions involves benthic classification and contribution, ocean environment, and biological webs.
    - Example: Understanding the low pH ocean involves biology, chemistry, physical transport
  - Many technical needs involved integrated projects/data
- General Strengths and Opportunities
  - Matched our own assessment (in SWOT)
  - Video, samples, benthic mapping highlighted
  - Data quantity (and context)

Community Needs: Visualization/Integration

- Way more feedback than any other topic
- First, let's talk about the process.
  - We'll use Quality Control as our example.

Community Needs: Quality Control Inputs

FC: Nothing stops us, just need more facilities. Just takes a lot longer to get work done. Where is biggest bang for our buck? FC: **If it's as easy to get data set as to google a person, it'd be done a lot more.** JB: Bottleneck isn't data or processing. Have to think about it. Process of discovery, mulling the data over, is very important. Lots of info can be in data you collect, not seen until you've looked at it a while. What tools do you use? JB: Highly variable.One thing that would be useful: **comparison of data he's looking at with previous or other similar data, requires catalogging system of what was done and where. He uses expedition DB a lot to get reference data. Could maybe use VARS more now.**

How much of spin-up on data is QC? Depends on data source. ADCP data pretty easy to check, other sensors like deep sea pH are horrible. FC: **Stable sensors will have occasional bad points. Developing automated QC for those are easy. Other data requires actual thought, experience, and judgment.Not at place yet where we could train systems to QC a lot of the data**. JB: Could **put different types of flags on it.** FC: can be useful on large data sets, algorithms can be developed there. Tim and he spend a few hours after every CTD cruise making results look reasonable.

Interested in oxygen profile for area, temperature range during month. Can do that with CTD profiles now? Not on the bottom sometimes.

Community Needs: Quality Control Inputs

- Quality control issues: drift, validation, volume. Two issues with data quality control: 1) Automating it (setting up rules, pretty simple set). 2) Human eyeballs ("this other rule we forgot to tell you about"). ISUS now QC'd automatically. Works because it's a complicated measurement, if you can't fit the array to a model you just toss it out. Temperature is harder. Lots of problems pop out when combining data sets. Survey result can be internally consistent, globally misplaced -- can't tell until combining. This is argument for combining lots of data sets. Lots of little things can be done to improve QC.
- Quality control of data is another huge issue. Even if we have an easy way for people to enter data, we must have QC mechanism to flag/identify bad data. Don't get rid of it! Just flag it! There was talk about peer voting on data quality (like Amazon's customer review). When people enter data, there should be some server mechanism that gives them an automated first cut at how their data looks. If the QC is bad, people won't use the data system. MBSystem was talked about as a good example here.
- for some data, not enough metadata (for example, are lights attracting the observation?). misidentifications are a problem
- Not at place yet where we could train systems to QC a lot of the data. Stable sensors will have occasional bad points. Developing automated QC for those are easy. Other data requires actual thought, experience, and judgment.
- put different types of flags on it.
- Too many false alarms. Our data sets are so sparse, we need to integrate before trying to detect. Need all of data together to cut down on false alarm rate.
- First event we need to be able to detect is the system not working. Once we have that working, we can move on to scientific events.

Community Needs: Quality Control

- Critical to making data reusable
- Automating "Easy QC" is feasible, but not "Hard QC" (the latter requires domain expertise)
- Integration (of data sets) key to effective QC
  - Can see outliers in bigger context, not by themselves
- Flagging data (not deleting it) should be standard practice
- Supporting active user participation (e.g., in flagging) would be valuable capability
- QC of <u>operations</u> should be first step
  - "Got data?"

Community Needs: Visualization and Integration

- Way more feedback than any other topic
- Lots of facets
  - 2D(map), 2.5D(GIS), 3D, 4D(time)…translate
  - Point/grid/track, models, images
  - Comparisons, context, overlays, icons, linkages
  - Data of interest: all, some, disparate, distributed
  - Searching, finding, integrating
- Visualization a recurring theme
  - Seeing data in context, at multiple scales, over time
  - Make it easy to use, integrative (multiple data sets)
  - Visualization valuable for both context and analysis

Community Needs: Data Organization

- Wish list
  - An easy way to “keep similar data together” when putting it into a repository
  - Highly usable interface—very important
  - Quickly access context for data
  - Replace notebooks (must be everywhere)
- Déjà vu
  - Many requests already exist in MBARI tools
  - May be an important question to pursue

Community Needs: Tools & Applications

- Many domain-specific requests
  - Bathymetry data management & visualization
  - Eddy correlation analytical tools
  - Enhance video mining tools: VARS, AVED
  - At-sea access to data collections
  - Extract quantitative data from non-transect dives
  - Real-time display with bathymetry, frame grabs
  - Generalize benthic analysis tools (even to 3D, time)
  - Generalize MDB (Scholin ESP) database features
  - Enhanced sample curation data management
- Often builds on existing strengths
- Often synthesizing patterns from existing data

Community Needs: Real-Time & Event Response

- Real-time developments downplayed
  - Usually fell into "would be nice" category
  - May be hard to envision revolutionary tools
- Event Response important in a few cases
  - First emphasis is on human interaction with at-sea systems

How should SDS proceed? (Advice from Science)

- Lots of input and specific recommendations:
- Marketing: early deliverable, make it easy to use
- Strategic Planning: look at the overall big picture
- Policies: data should go out (unless it shouldn’t)
- Priorities: users vs products, new vs upgrades
- Data Mgt: (don’t) centralize data
- Data Security: Very important, make consistent across distributed systems
- Best Practices: include users (as testers), create products (with real data), stay close to ‘raw data’

How do we proceed?

- Science input collected and synthesized
  - More input welcome at any time (now is good)
  - Science understanding and acceptance of our results will be key to our success
  - Science strategic planning will contribute further to our understanding
- Next Up: Integration with our SWOT* analysis

* SWOT=Strengths—Weaknesses—Opportunities—Threats

Results of SWOT

- Process
- Strengths
- Weaknesses
- Opportunities
- Threats
- Strategy

Results of SWOT

- Process
  - SWOT is initial step in strategic planning
  - Helps us focus on what we can actually control
  - Formed into four groups to focus on each:
    - Strengths – John Graybeal, Mike Godin, Dan Davis
    - Weaknesses – Dan Davis, Duane Edgington, Kevin Gomes
    - Opportunities – Dave Caress, Dan Davis, John Graybeal
    - Threats – Kevin Gomes, Mike Godin, Duane Edgington

Results of SWOT -- Strengths

- Strengths
  - People with Experience
  - Diverse Data Sets
  - Environment – People Interactions
  - Environment – Policies/Practices
  - Environment – Education/Outreach
  - Environment – Stable Resources
  - External Connections
  - Very Cost Effective

Results of SWOT -- Weaknesses

- Weaknesses
  - Limited Resources
  - Lack of consensus and focus on what is important
  - Inconsistent Engineering Process

Results of SWOT -- Opportunities

- Opportunities
  - Leadership
  - Observatory Data Logging and Archiving
  - Integrating Modeling with Observations
  - Integrating Diverse Data Types into Useful Data Products
  - Data Mining and Query Algorithms

Results of SWOT -- More Opportunities

- Integrating Data Archives from Multiple Institutions
- Developing Standard Interfaces for Interoperability
- Developing Quality Control
- Understanding Optimum Sampling Scales in Time and Space
- Automated Data Interpretation
- Virtual Observatories ("Point-and-click Oceanography")

Results of SWOT -- Threats

- Threats
  - Competition and Barriers to Collaboration
  - Disconnection from Evolution of Community/Industry Standards and Technology
  - Both the Instability of, and our Inability to Leverage, external Resources
  - Risks in Sharing MBARI's Data and Technology

Status of Planning Process

- Strategic Planning Points
  - Focus on Solutions for Technology-Limited Marine Science
  - Insist on Teamwork
  - Insure Optimum Use of our Assets and Capabilities
  - Complement Rather Than Duplicate
  - Demonstrate the Worth of our Efforts
  - Provide Effective Communication, Cooperation and Collaboration
  - Seek External Advice on the Quality of our Work
  - Expand our Relationships with Regional Institutions
  - Transfer our Successes to the Marine Research Community
  - Communicate with the Public

Possible Directions

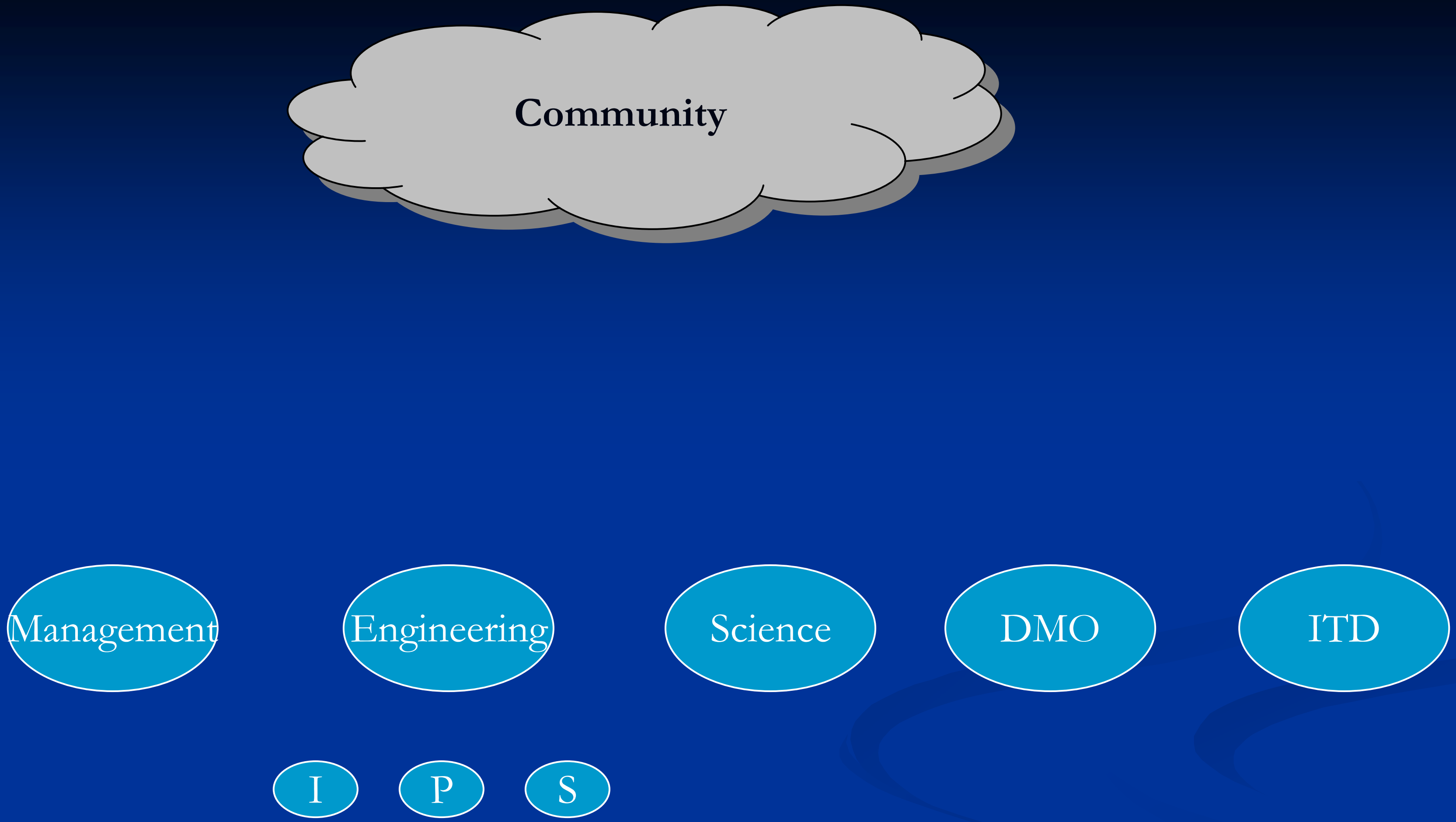
- Provide/Develop/Integrate Tools for Query, Integration, Visualization (4D) and Modeling
- Enable Data Understanding
- Enable Data Discovery
- Provide Mechanism for Data/Metadata Storage
- Advanced Ocean Observatory Functions (Event Response, Adaptive Sampling, Lab-in-a-can)
- Sample Curation

Issues

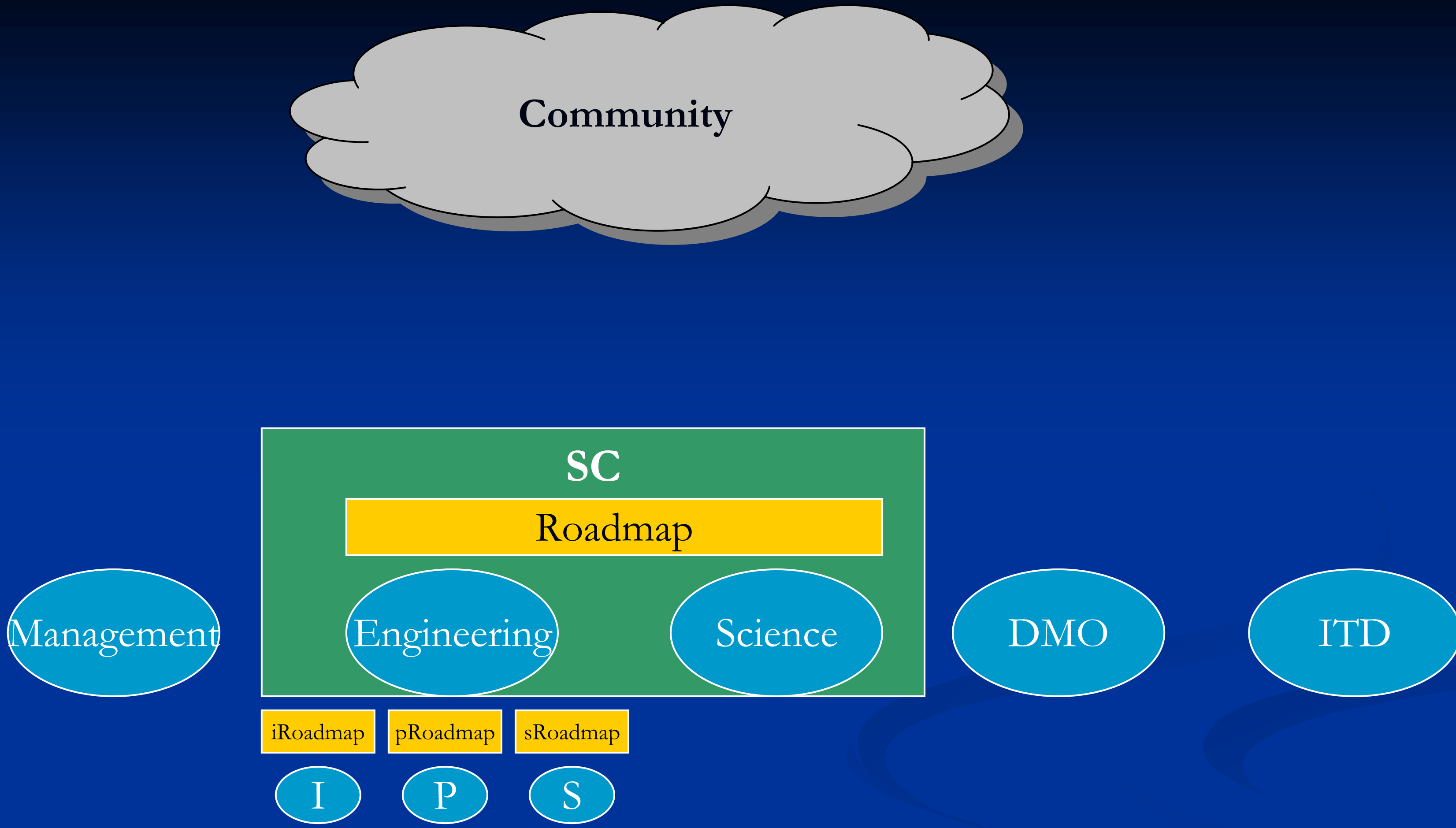
- Heavier inter-group dependencies
  - A trend in Oceanography is towards integrated systems and tools that work in those integrated systems
- We risk fragmentation
  - With whom?
    - Within Engineering
    - Between Science and Engineering
    - With the Community
  - How?
    - Problem Domain
    - Solution
    - Duplication of Work
    - Expectations e.g. visualization

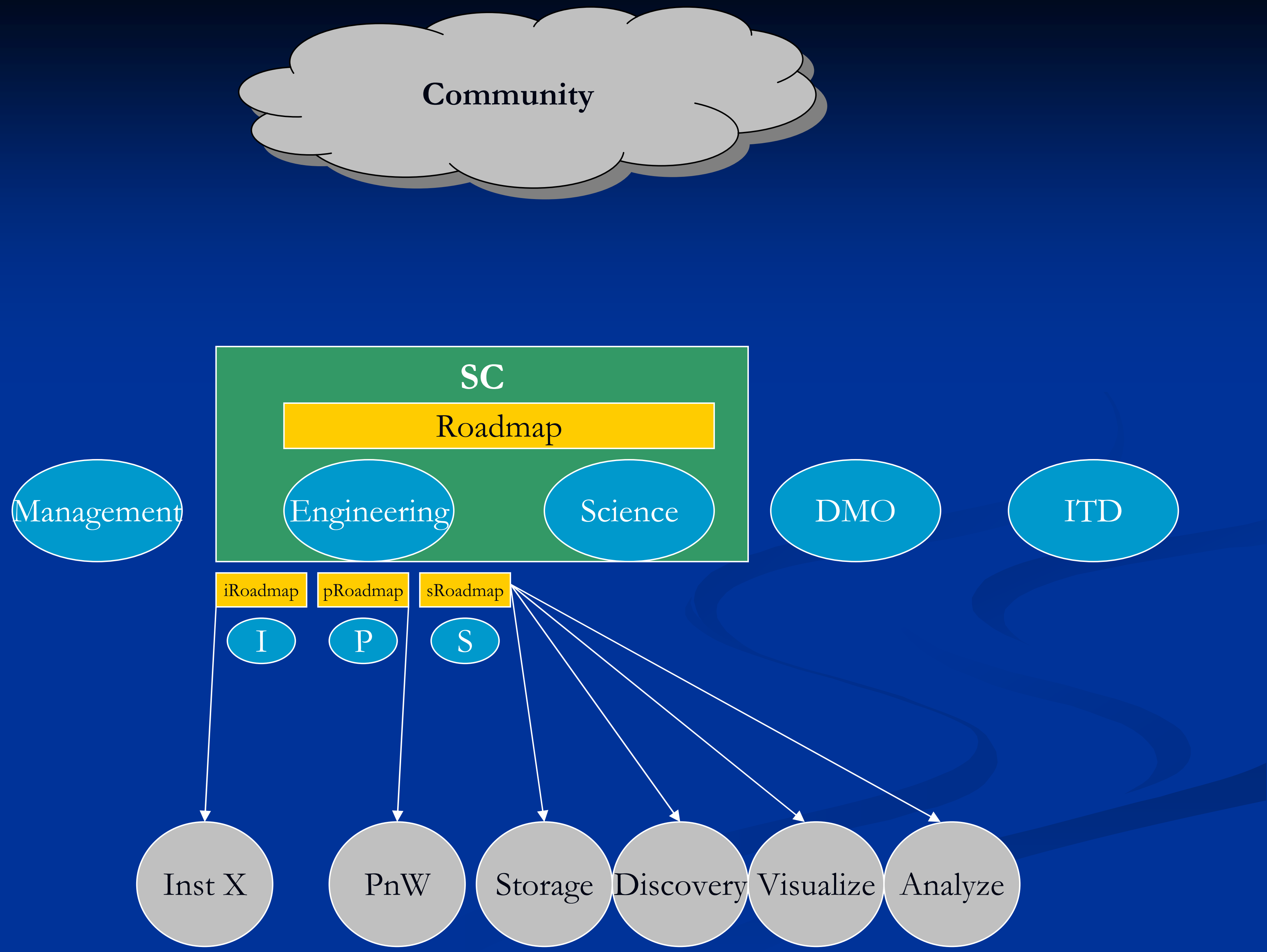
What do we need?

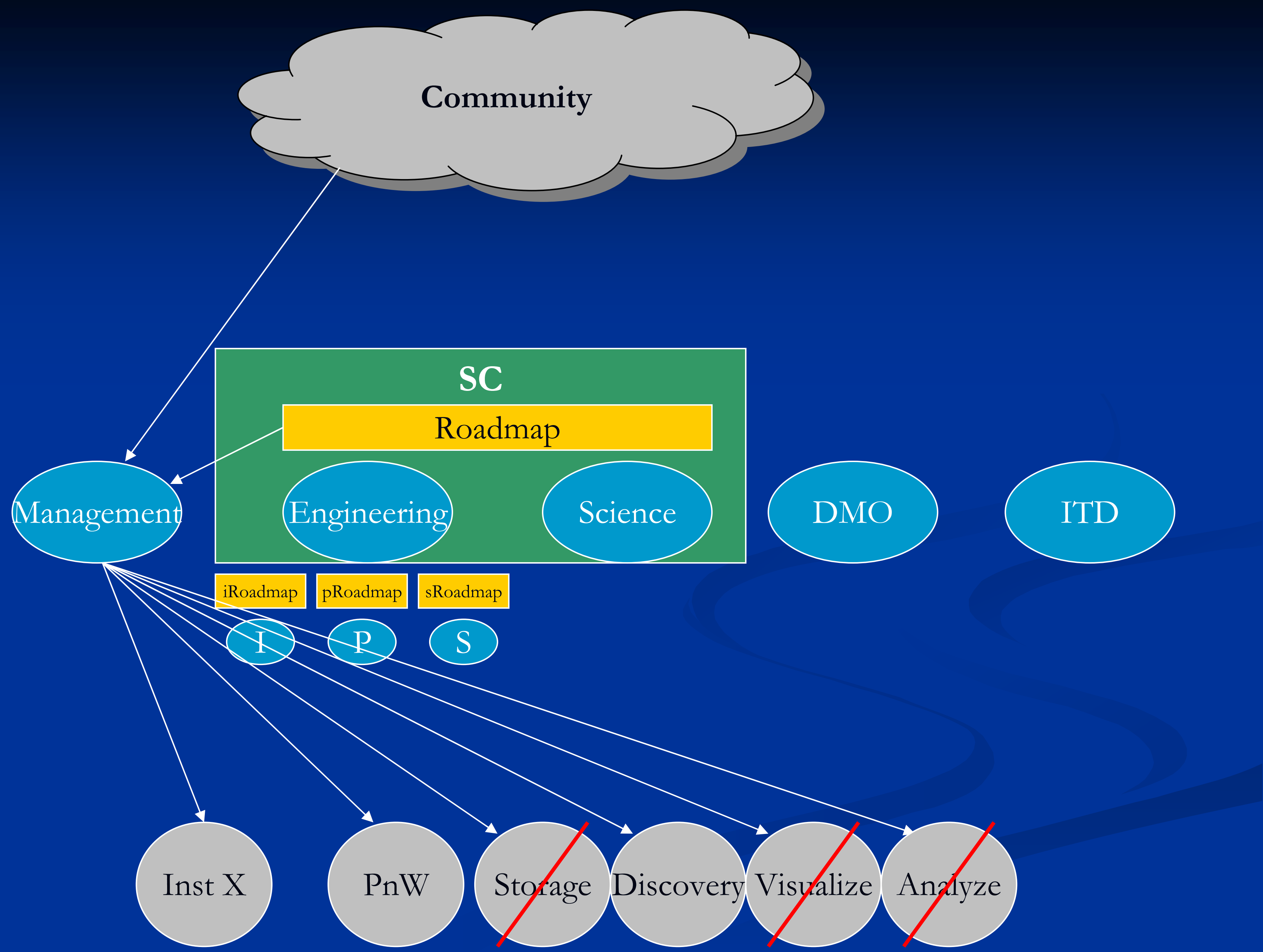
- A group like Strategic Council/Systems Engineering (SC/SE) that can help drive:
  - A roadmap for each subgroup and one that for all of engineering that drives a common focus
  - A process for re-synching focus and roadmap.
  - The process and authority to make the hard decisions that will allow our small engineering org to be successful
  - Standardization of instruments/platforms
- Means of defining and enforcing necessary policies
- Metadata and processes to capture it

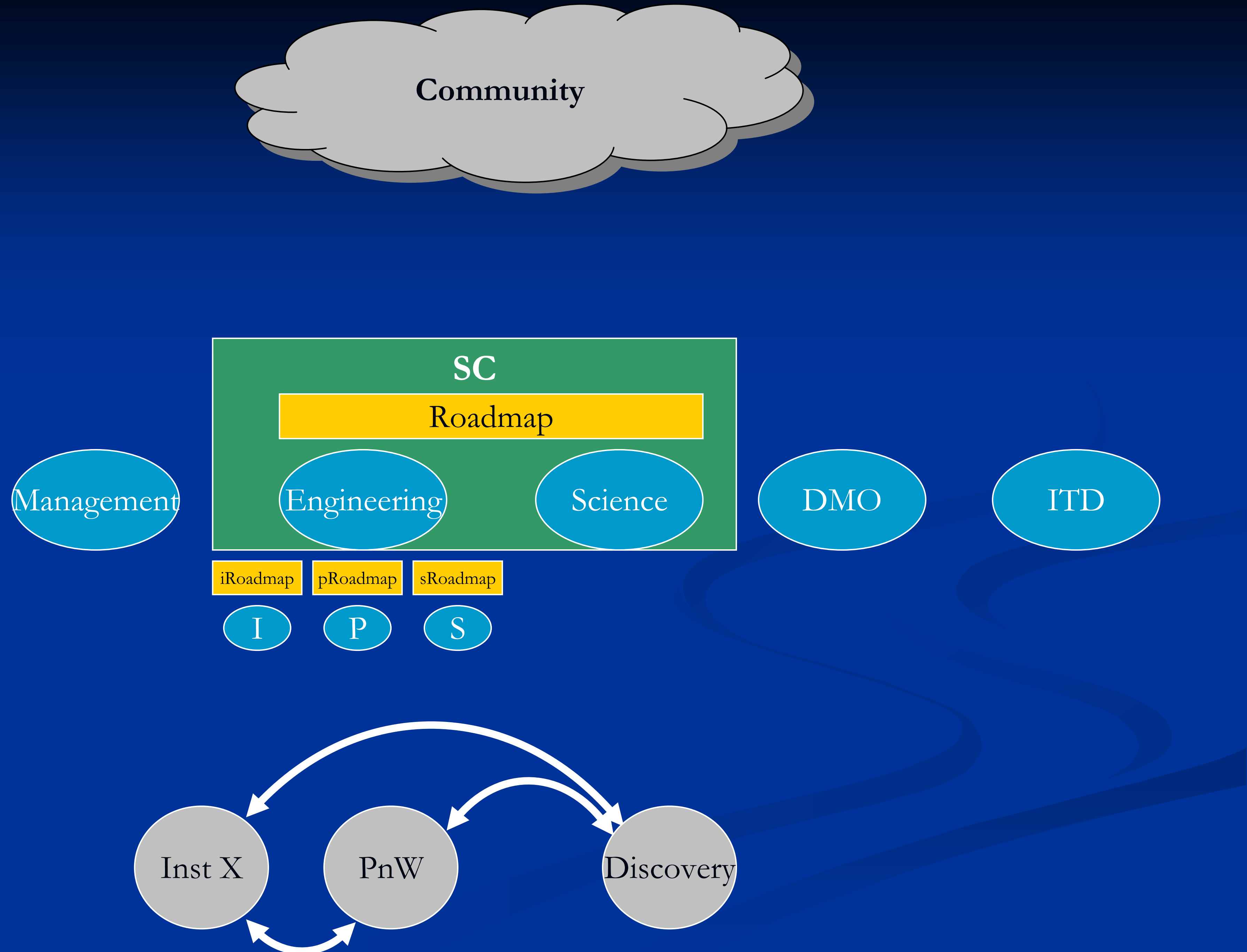


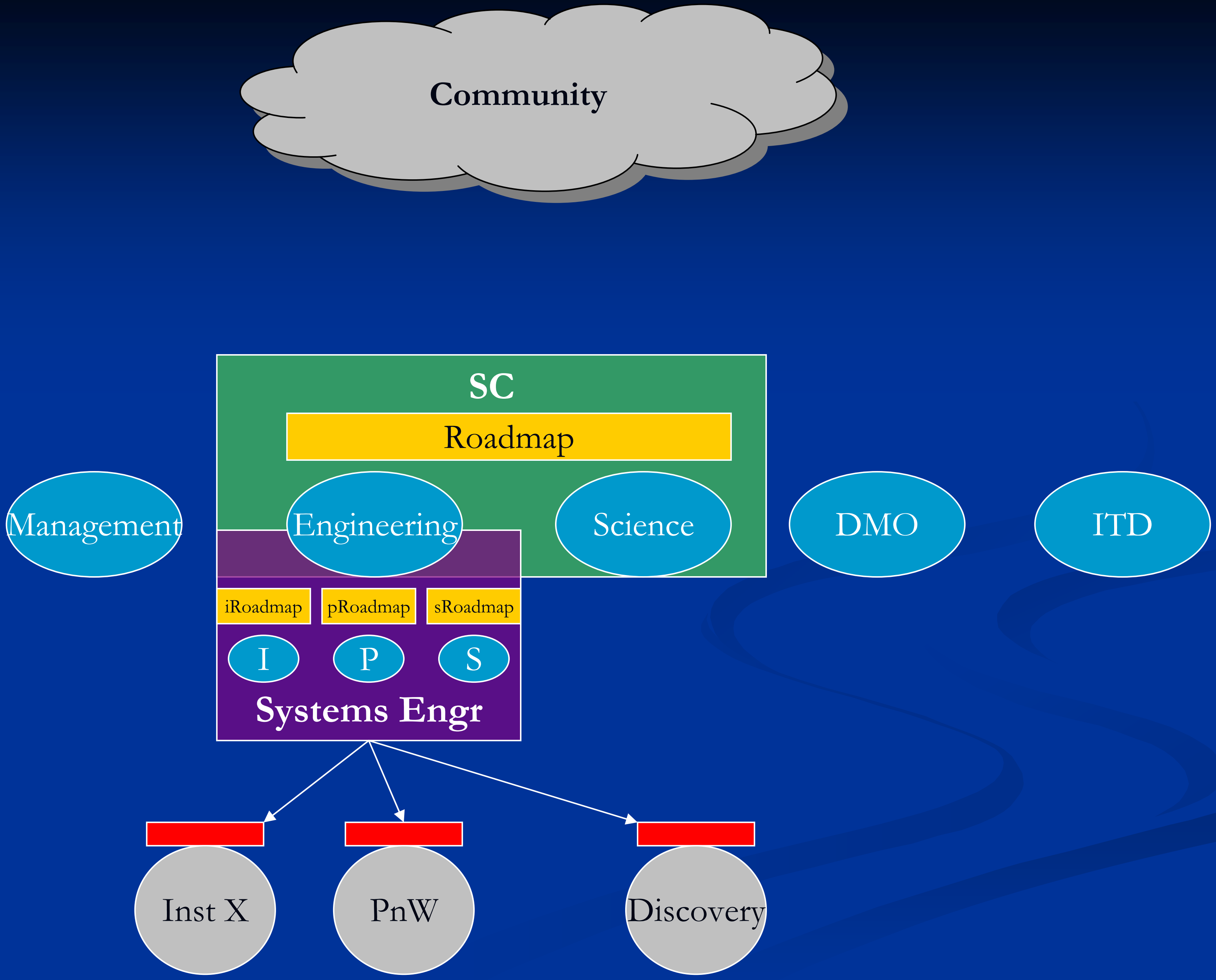
Community
Management
Engineering
Science
DMO
ITD
I
P
S

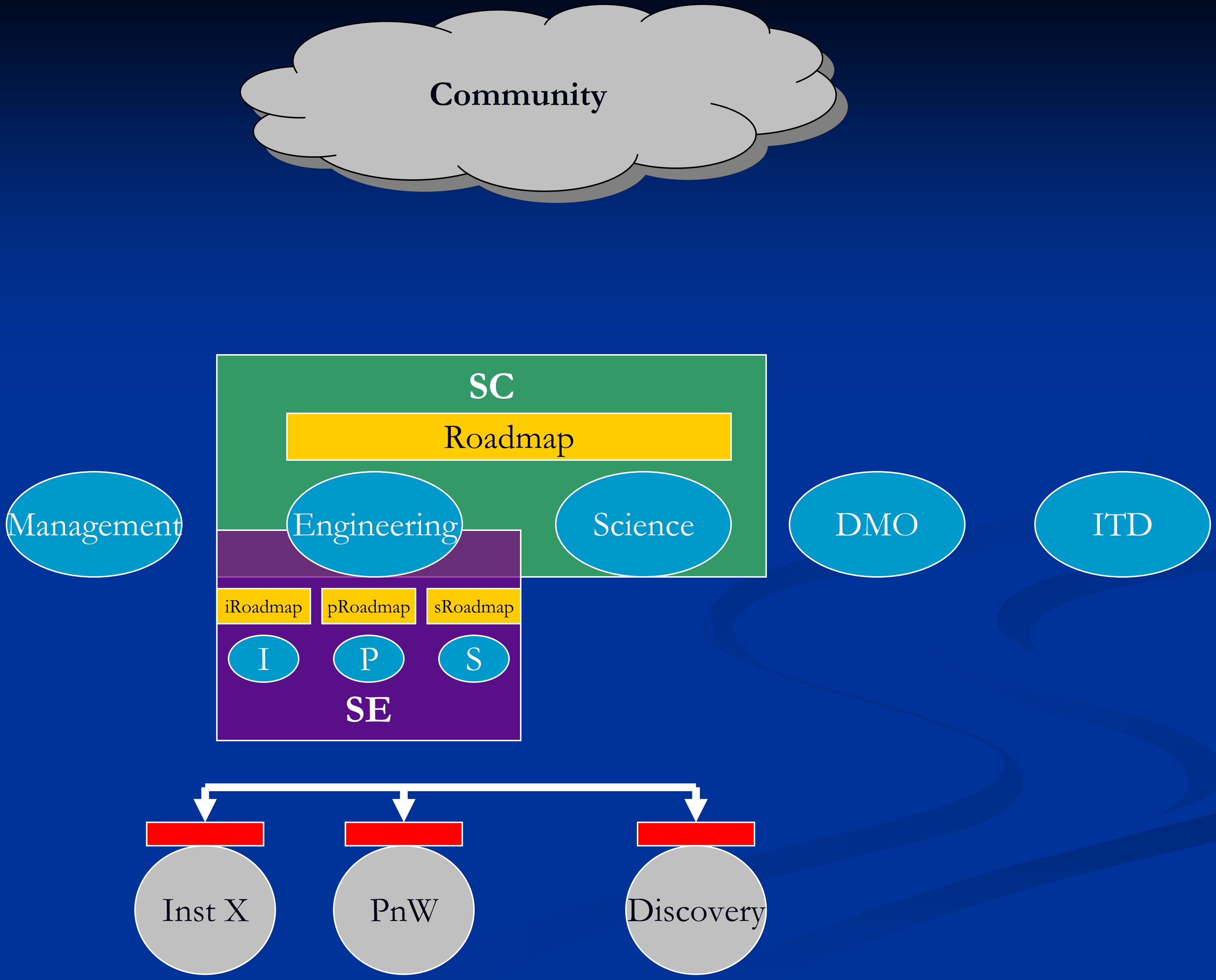

Community
SC
Roadmap
Management
Engineering
Science
DMO
ITD
iRoadmap
pRoadmap
sRoadmap
I
P
S

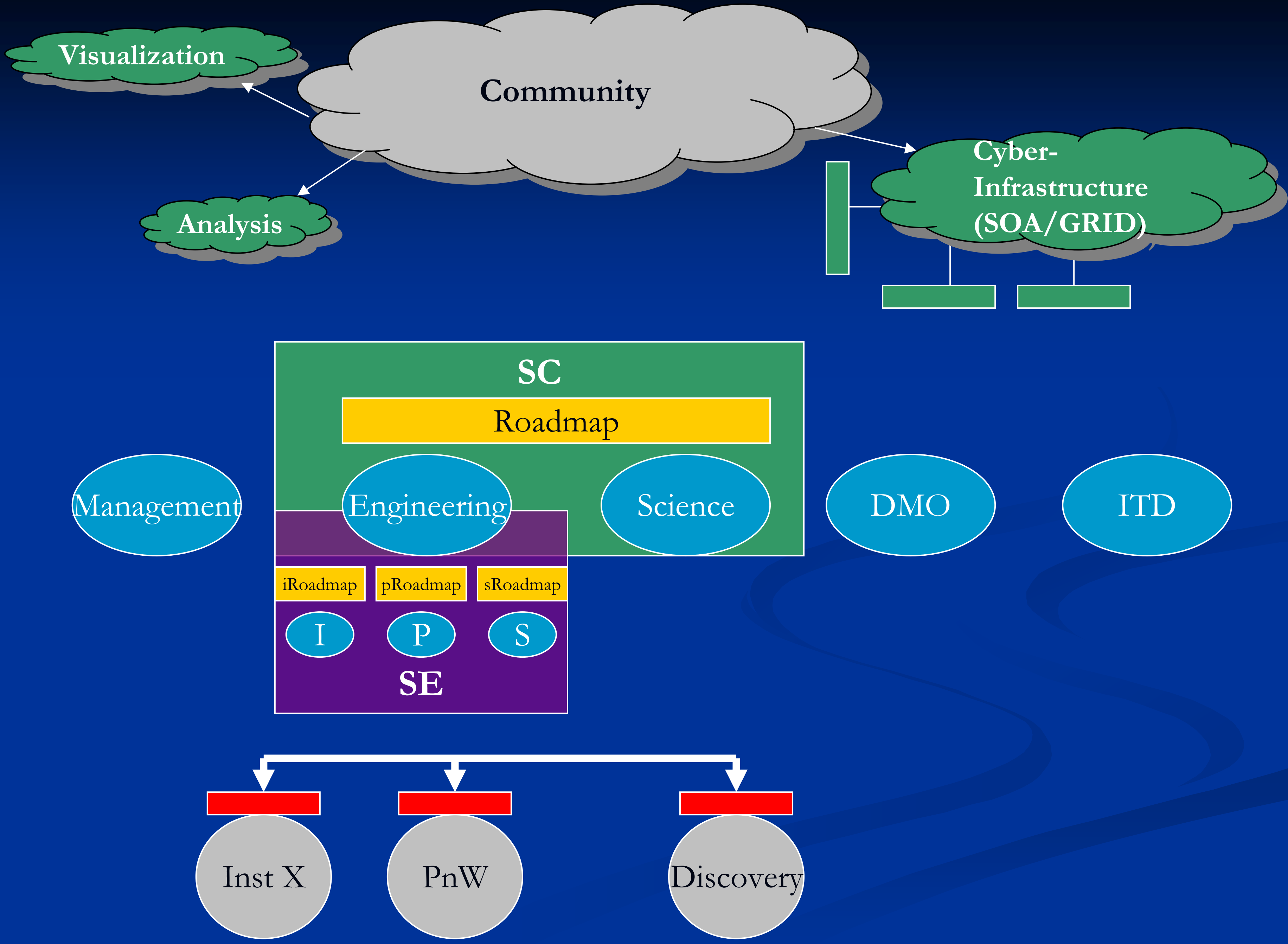

Community
SC
Roadmap
Management
Engineering
Science
DMO
ITD
iRoadmap
pRoadmap
sRoadmap
I
P
S
Inst X
PnW
Storage
Discovery
Visualize
Analyze

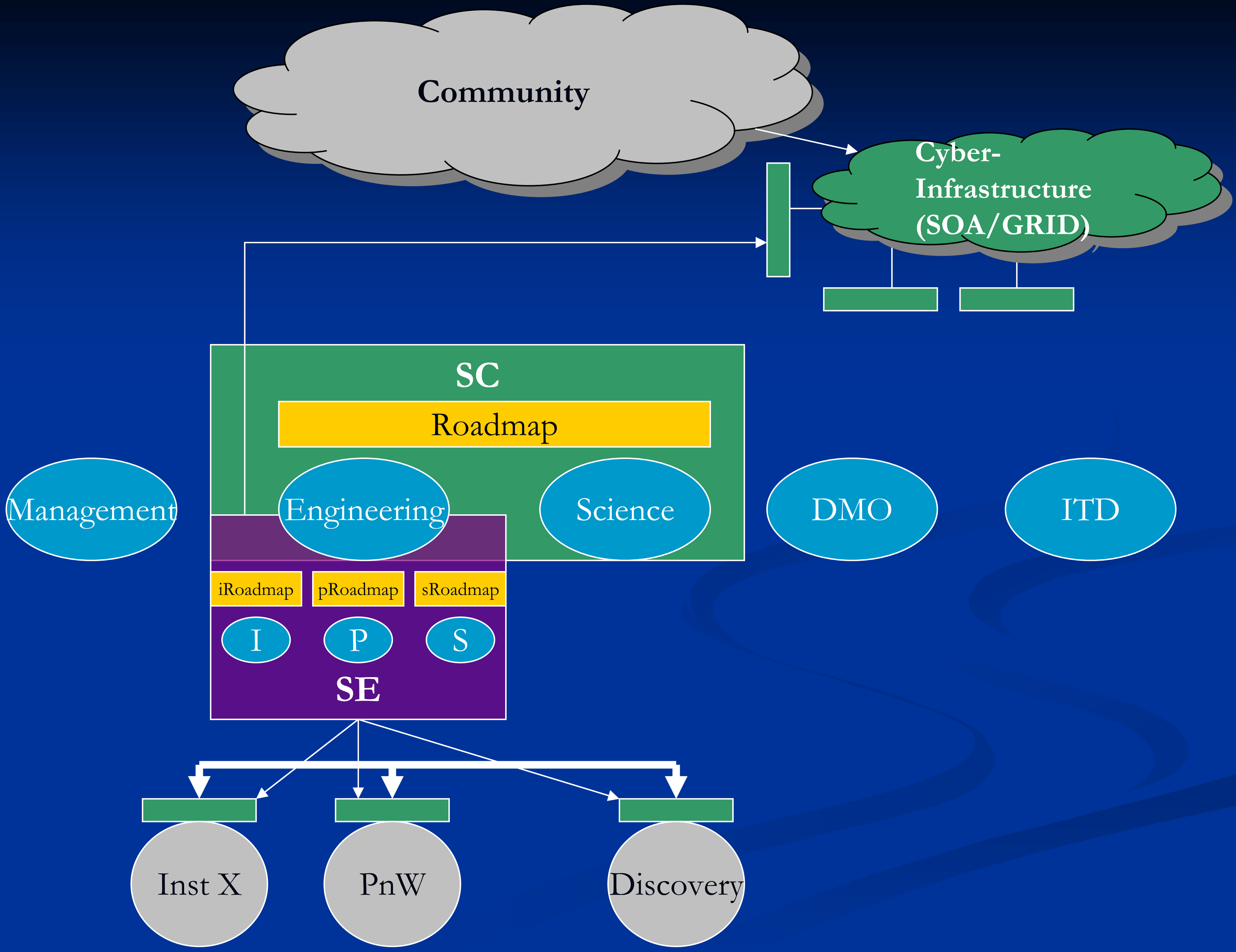

Community
SC
Roadmap
Management
Engineering
Science
DMO
ITD
iRoadmap
pRoadmap
sRoadmap
I
P
S
Inst X
PnW
~~Storage~~
Discovery
~~Visualize~~
~~Analyze~~

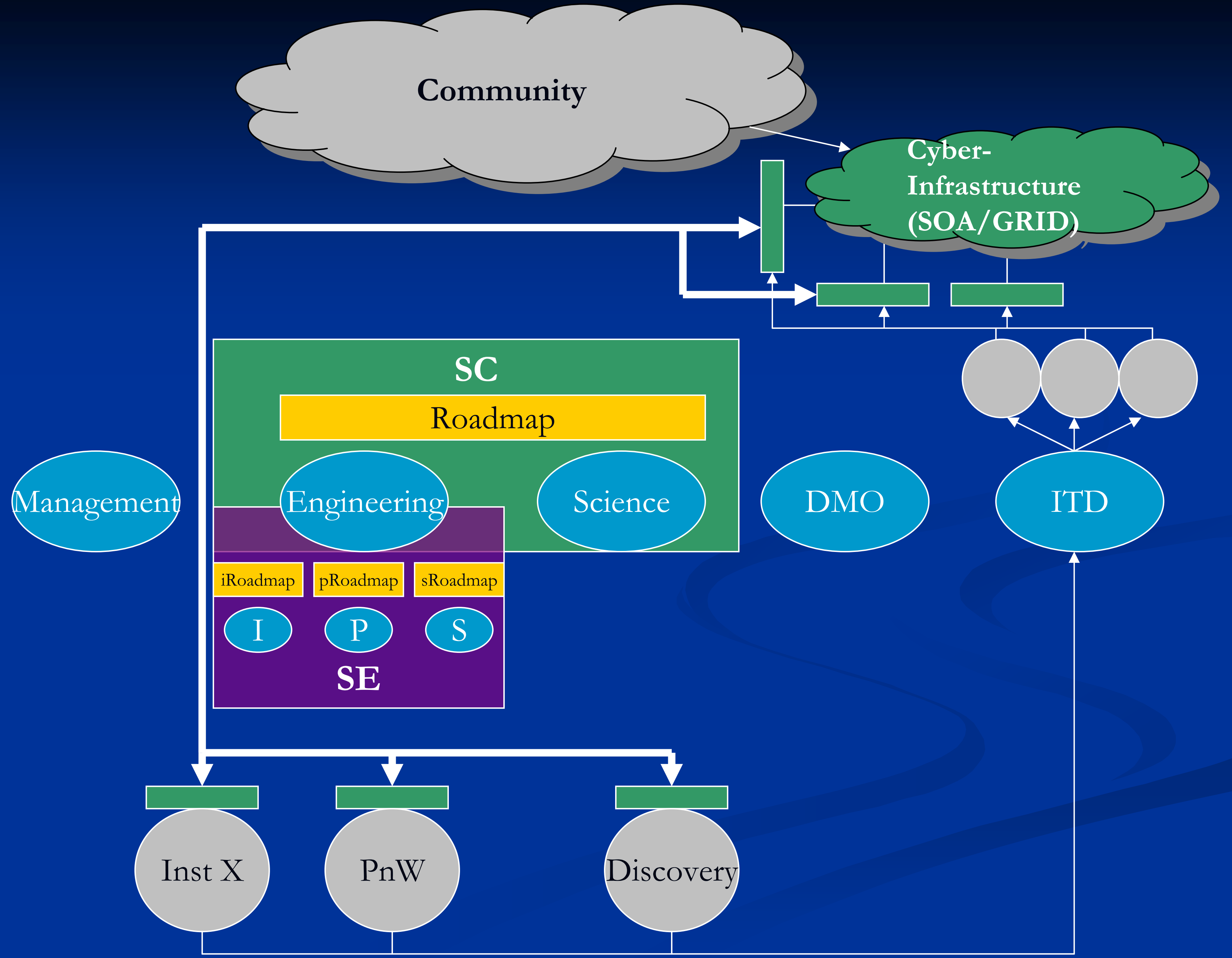

Community
SC
Roadmap
Management
Engineering
Science
DMO
ITD
iRoadmap
pRoadmap
sRoadmap
I
P
S
Inst X
PnW
Discovery


Community
SC
Roadmap
Management
Engineering
Science
DMO
ITD
iRoadmap
pRoadmap
sRoadmap
I
P
S
Systems Engr
Inst X
PnW
Discovery

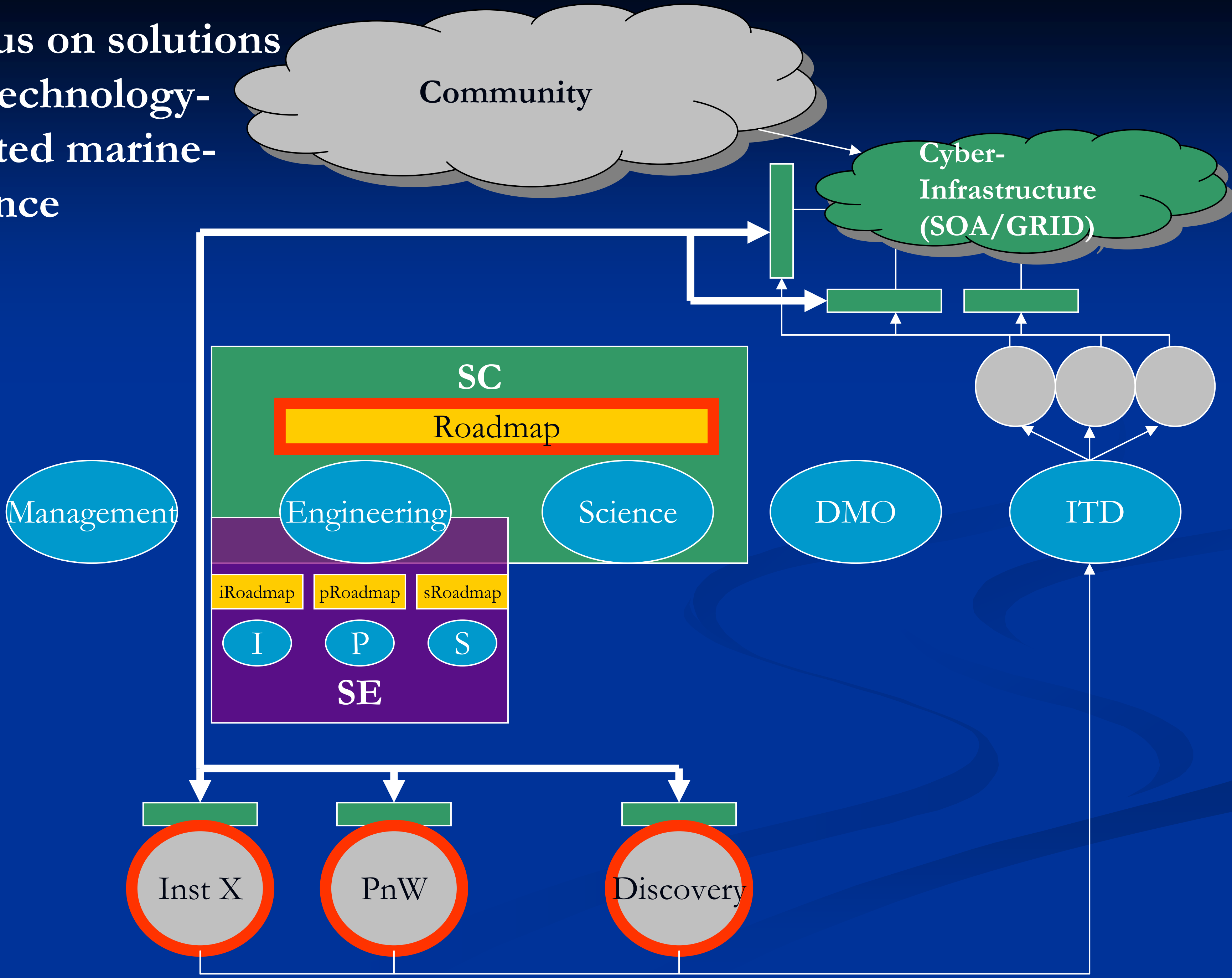

Community
SC
Roadmap
Management
Engineering
Science
DMO
ITD
iRoadmap
pRoadmap
sRoadmap
I
P
S
SE
Inst X
PnW
Discovery


Community
Visualization
Analysis
Cyber-
Infrastructure
(SOA/GRID)
SC
Roadmap
Management
Engineering
Science
DMO
ITD
iRoadmap
pRoadmap
sRoadmap
I
P
S
SE
Inst X
PnW
Discovery

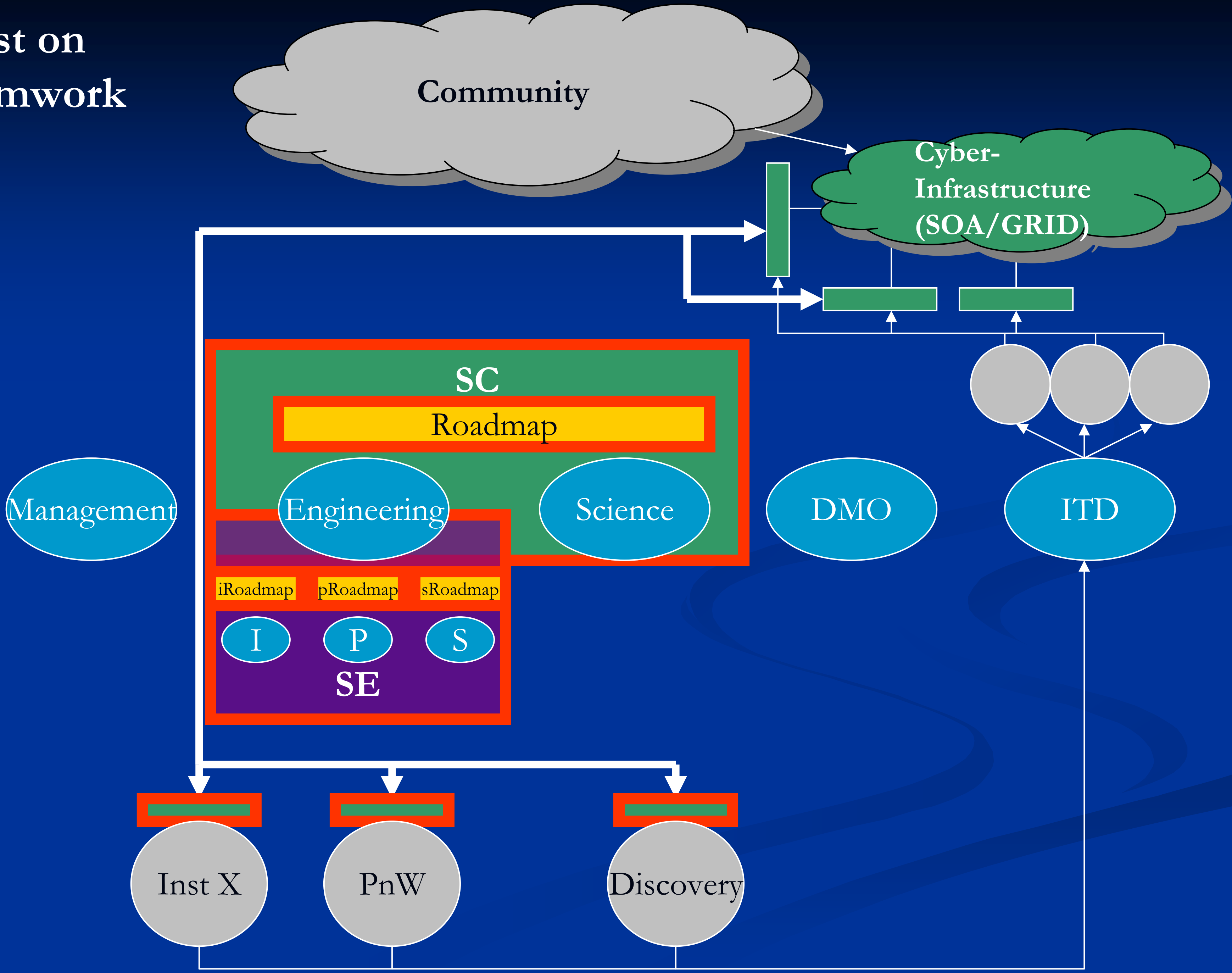

Community
Cyber-
Infrastructure
(SOA/GRID)
SC
Roadmap
Management
Engineering
Science
DMO
ITD
iRoadmap
pRoadmap
sRoadmap
I
P
S
SE
Inst X
PnW
Discovery


Community
Cyber-
Infrastructure
(SOA/GRID)
SC
Roadmap
Management
Engineering
Science
DMO
ITD
iRoadmap
pRoadmap
sRoadmap
I
P
S
SE
Inst X
PnW
Discovery

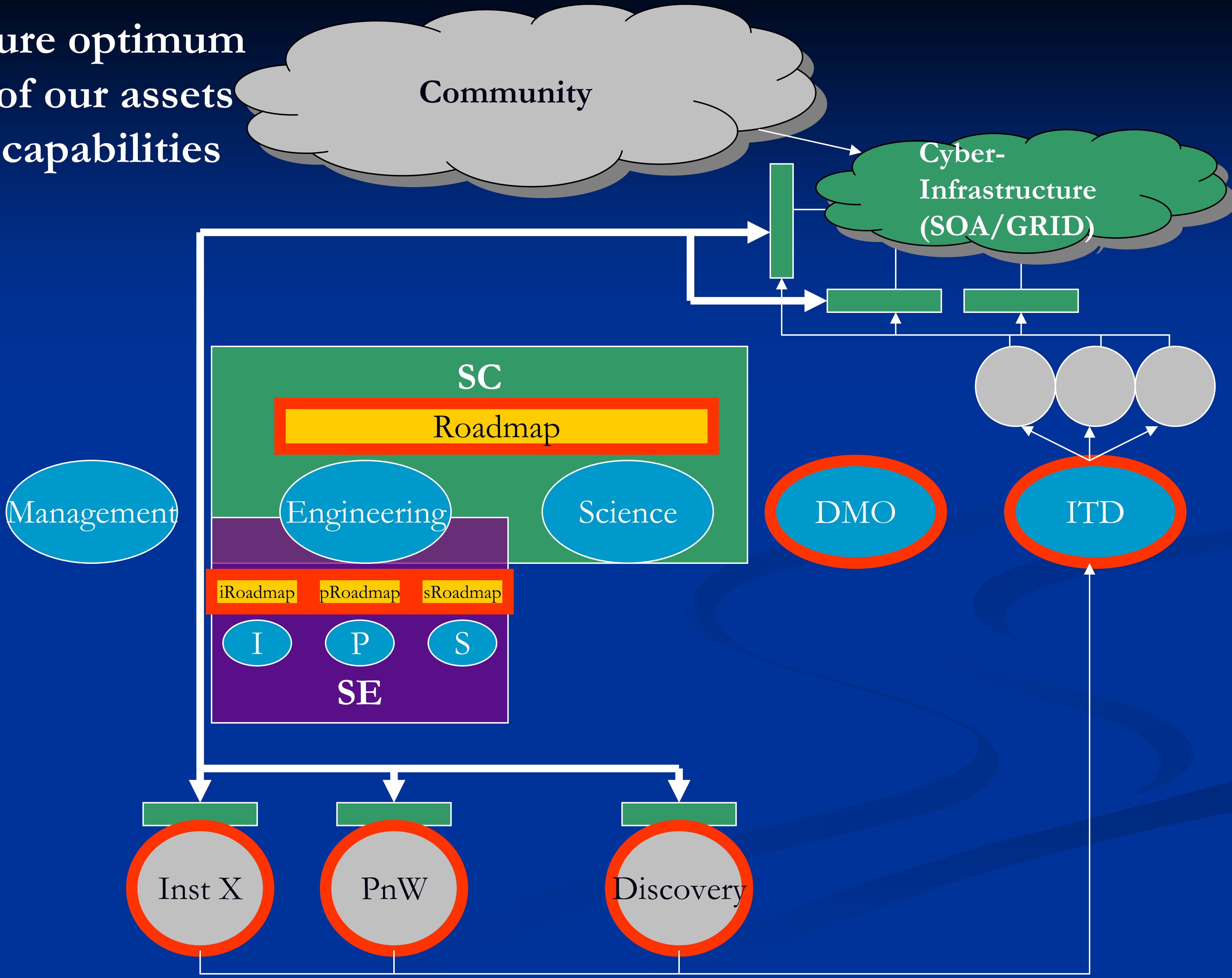
Focus on solutions for technology-limited marine-science



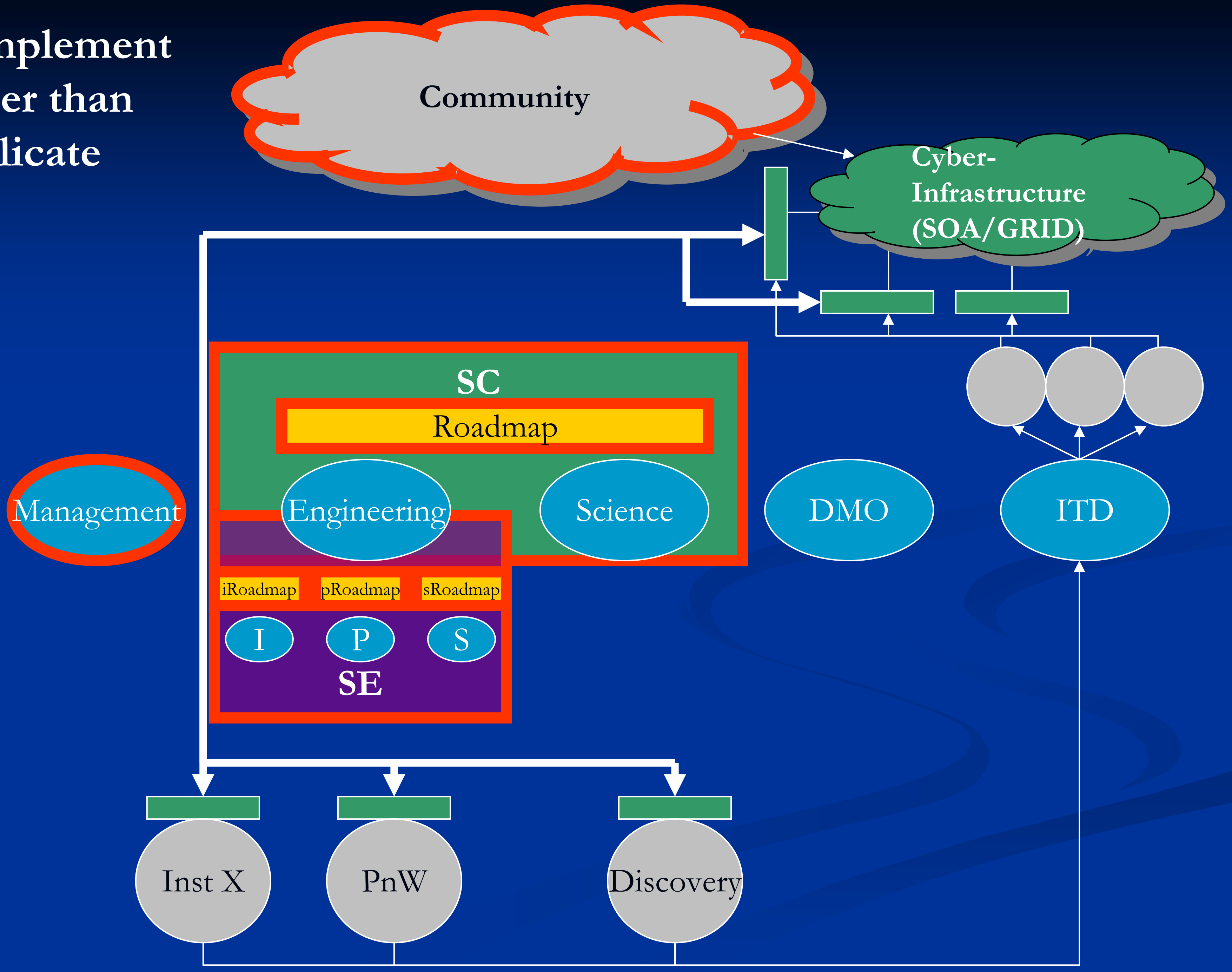
Insist on Teamwork



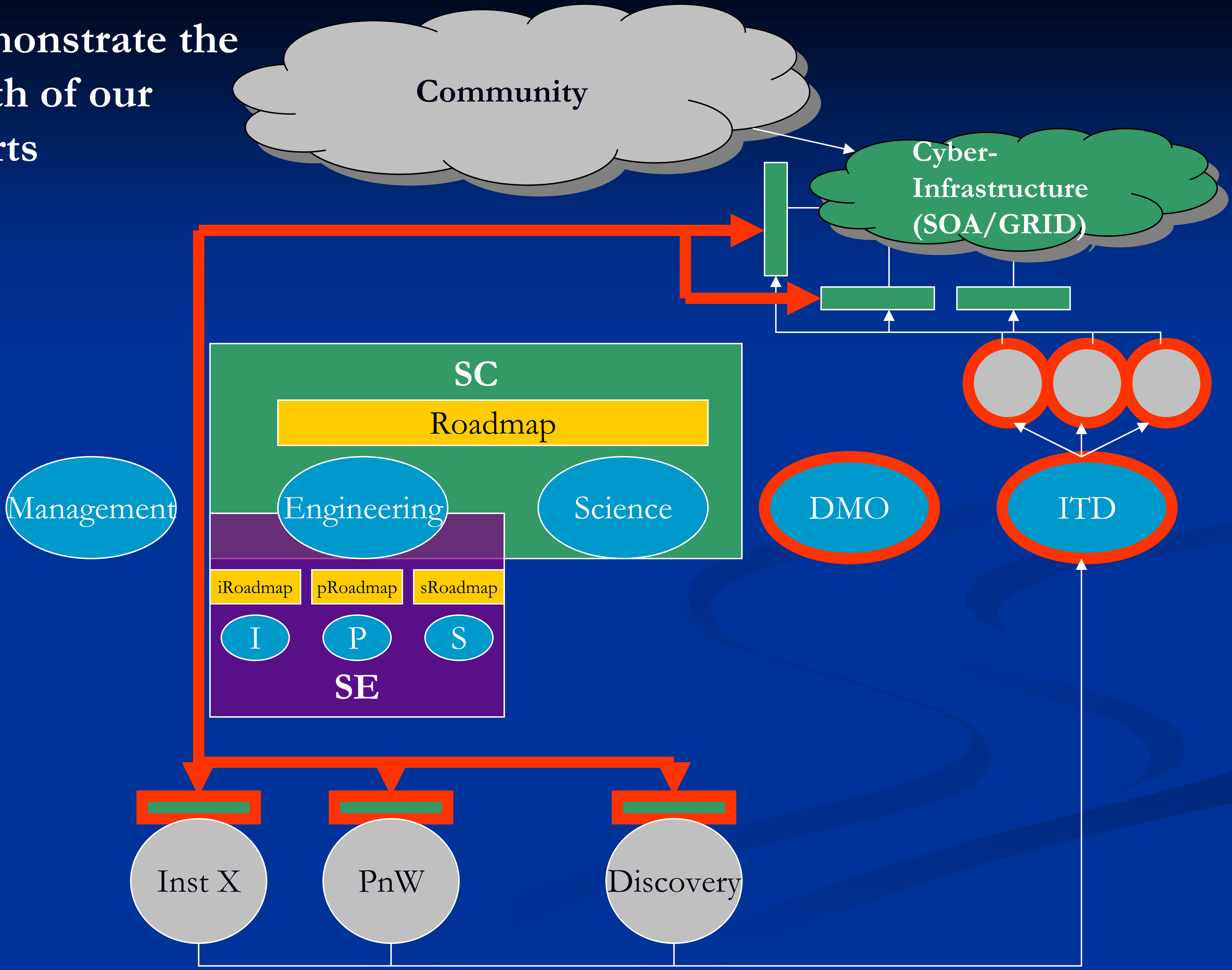
Ensure optimum use of our assets and capabilities



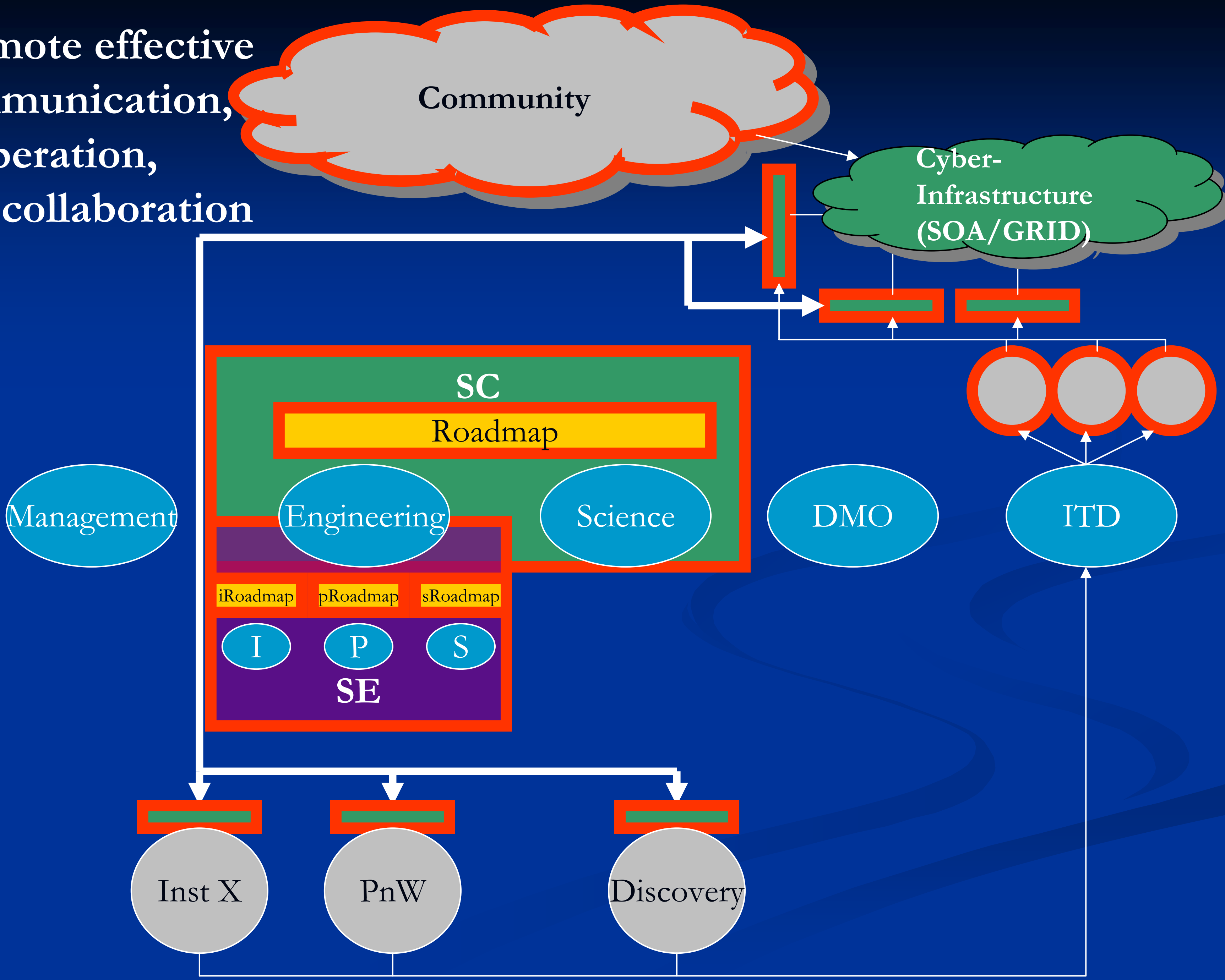
Complement rather than duplicate



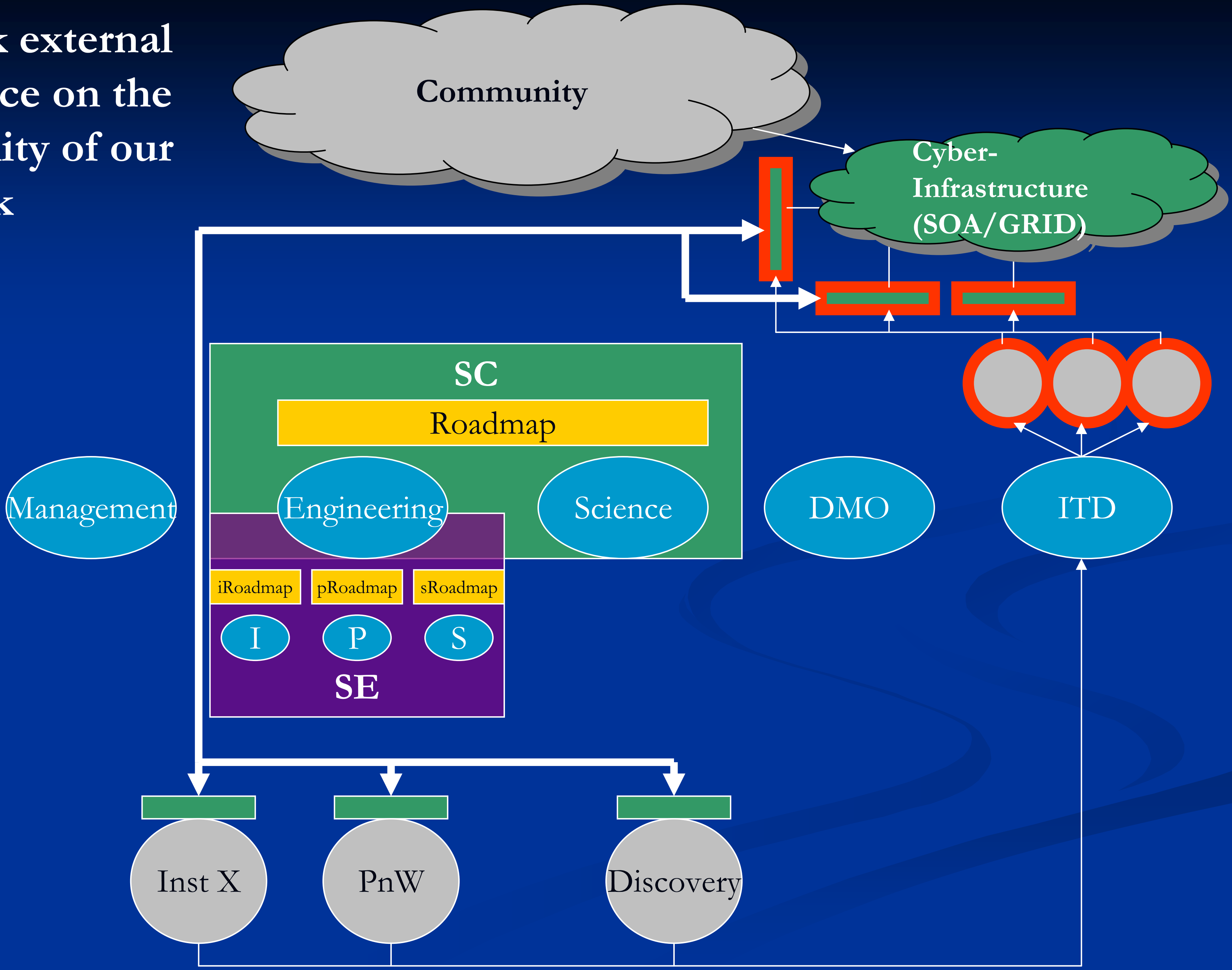
Demonstrate the worth of our efforts



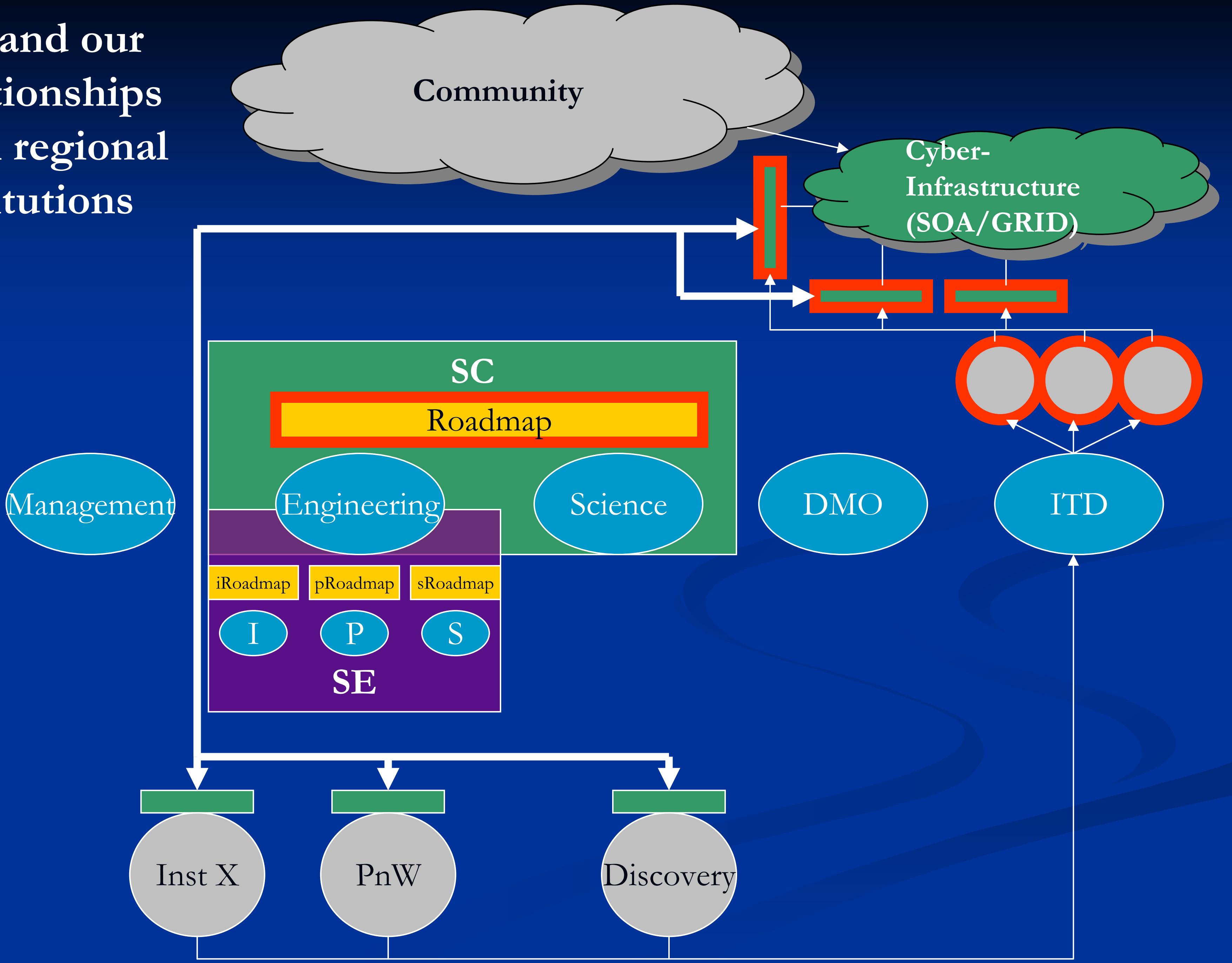
Promote effective communication, cooperation, and collaboration



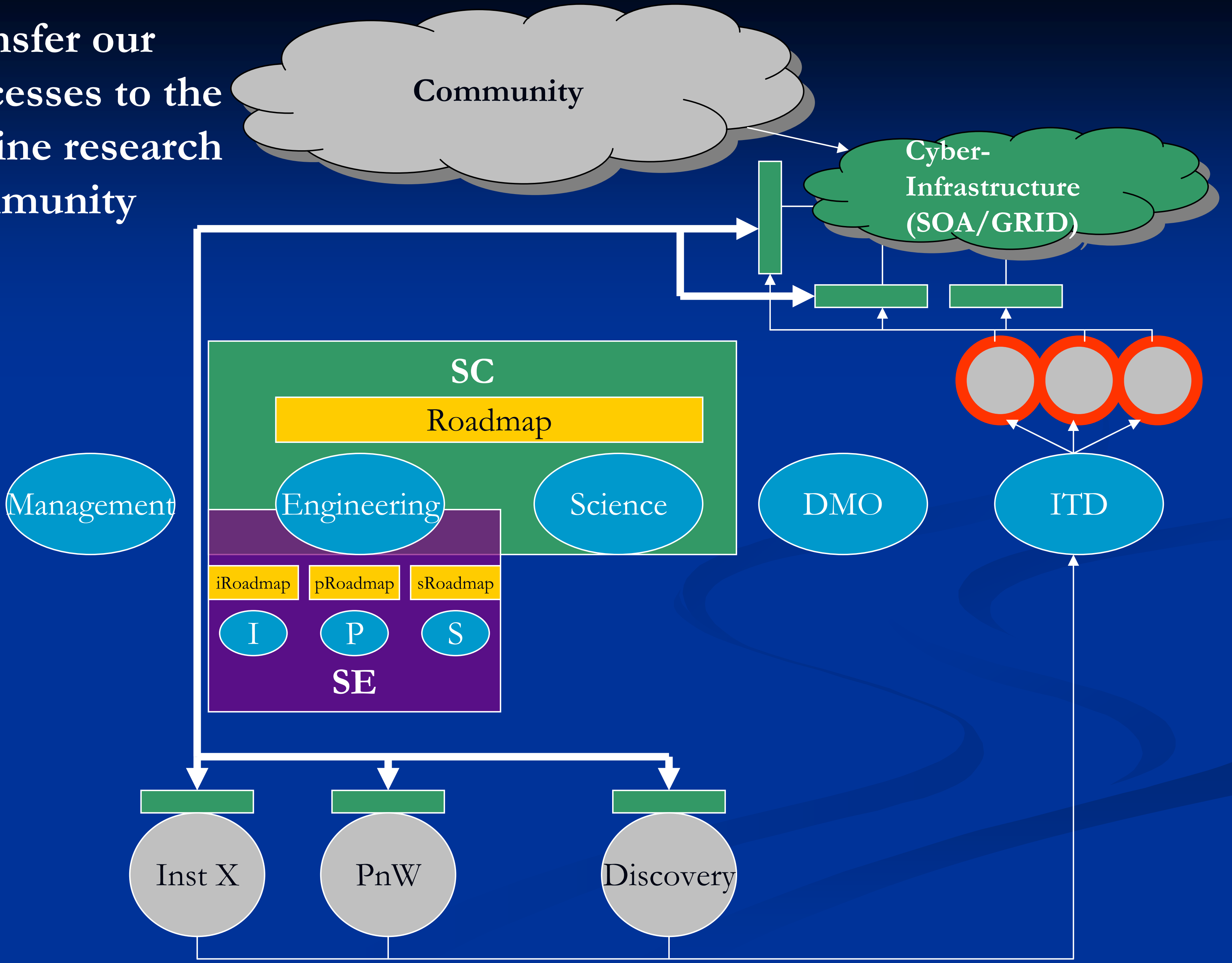
Seek external advice on the quality of our work



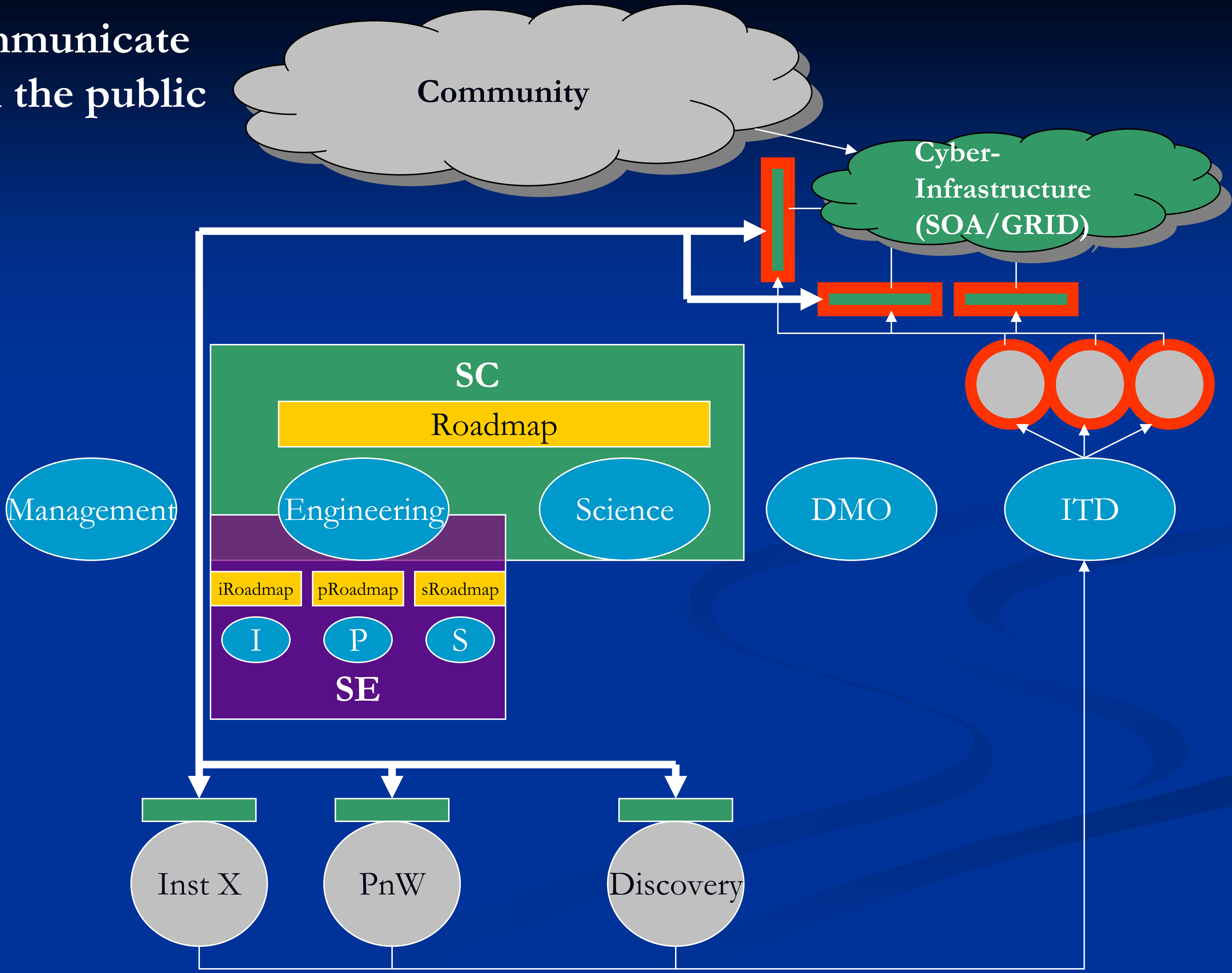
Expand our relationships with regional institutions



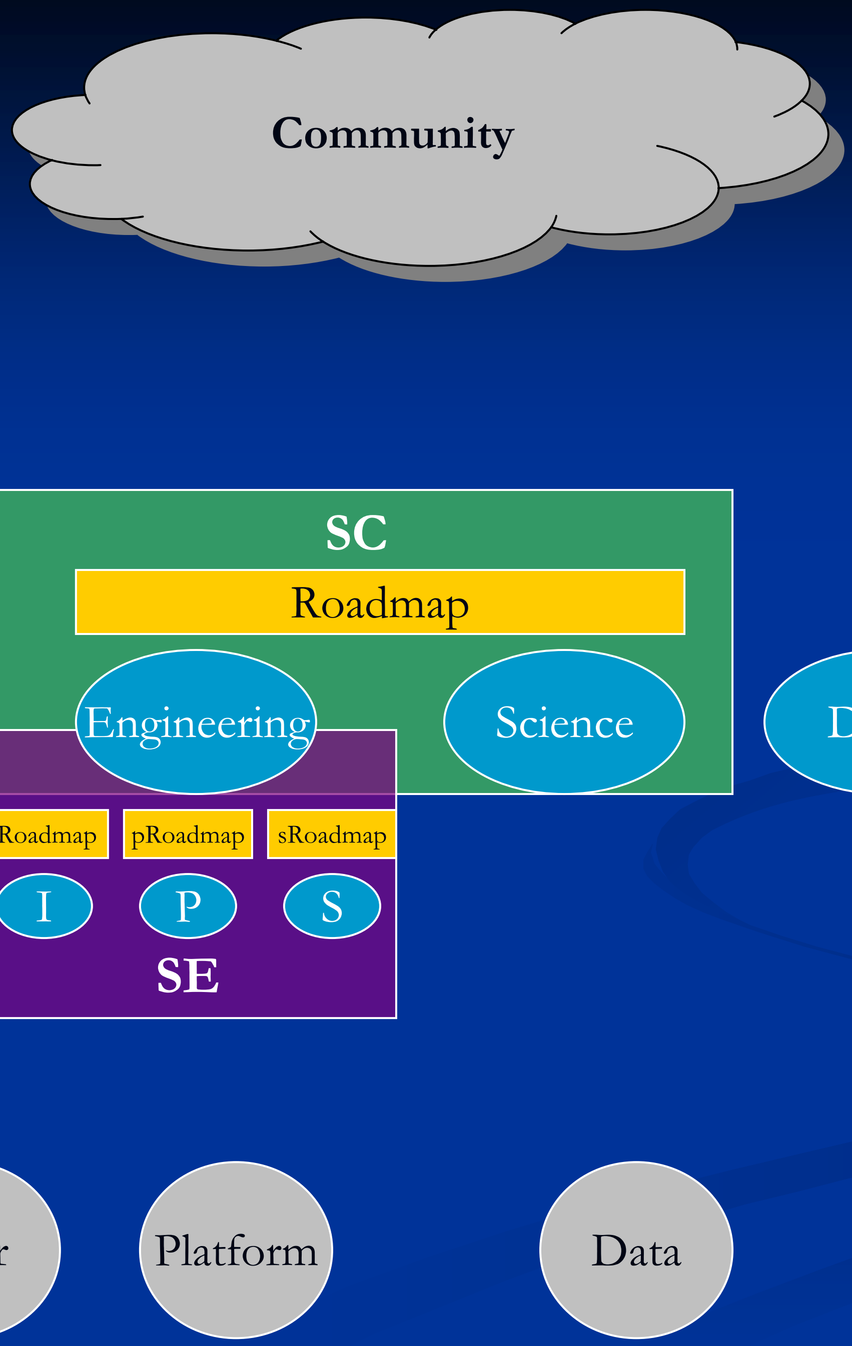
Transfer our successes to the marine research community



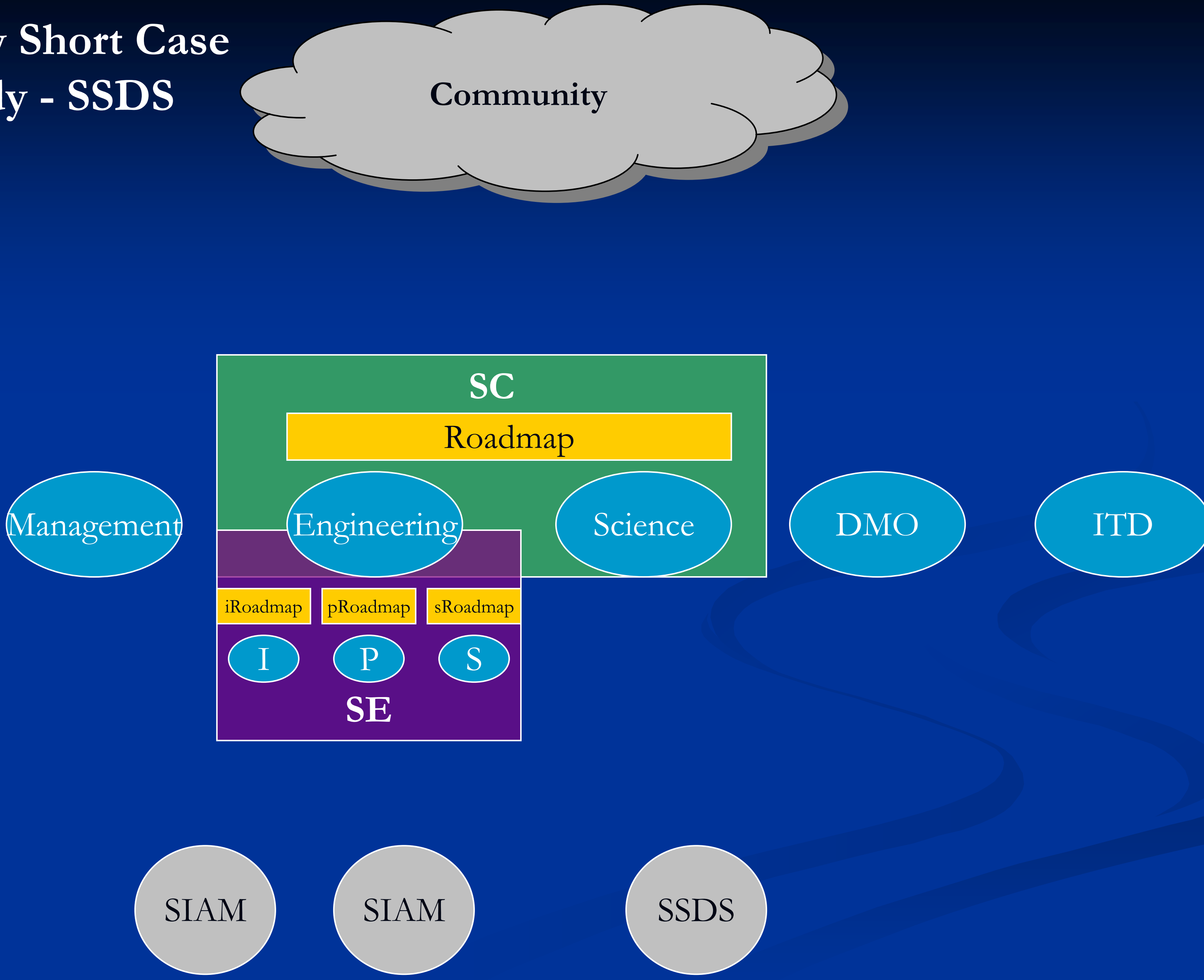
Communicate with the public



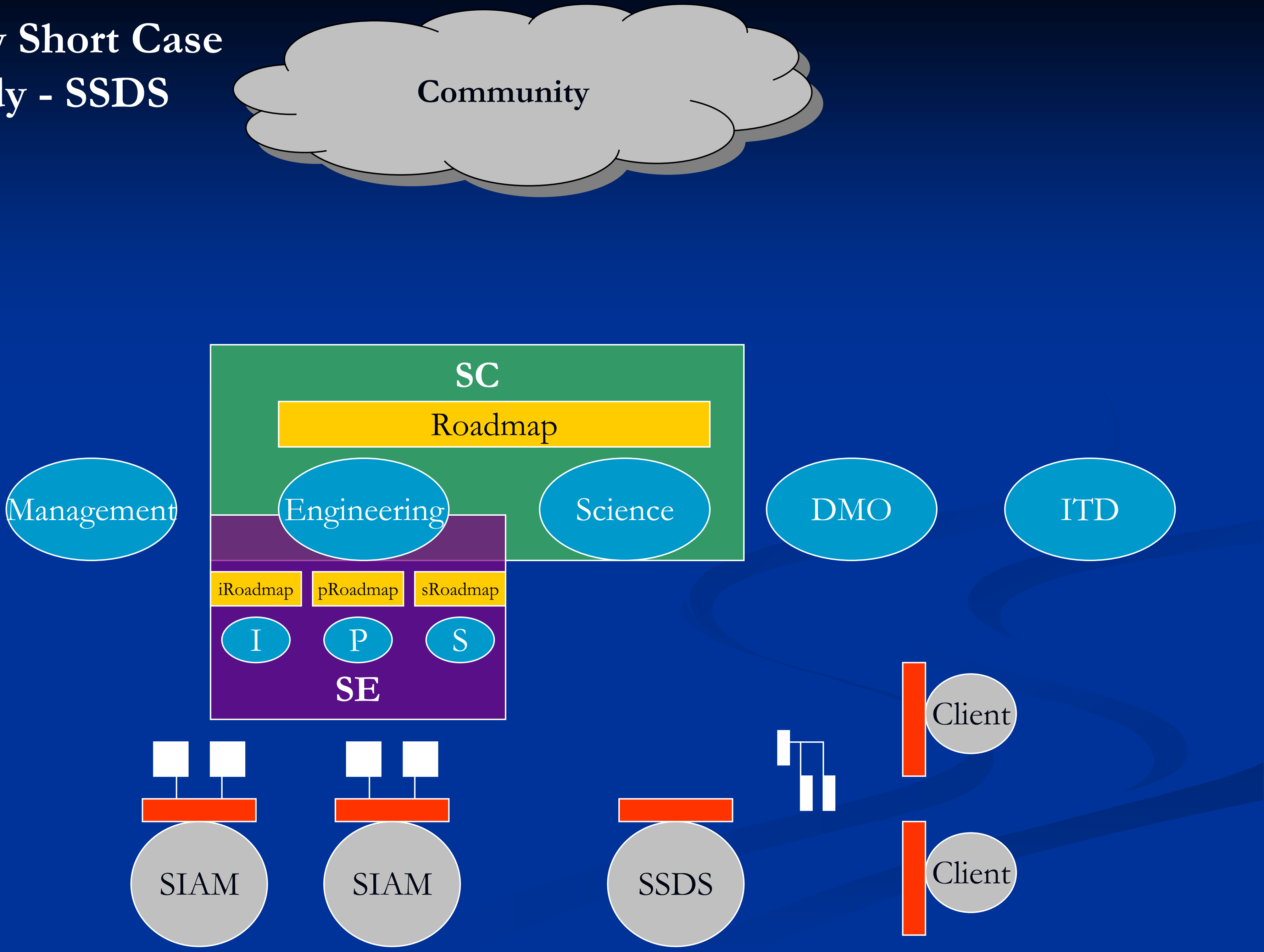
Very Short Case Study - SSDS



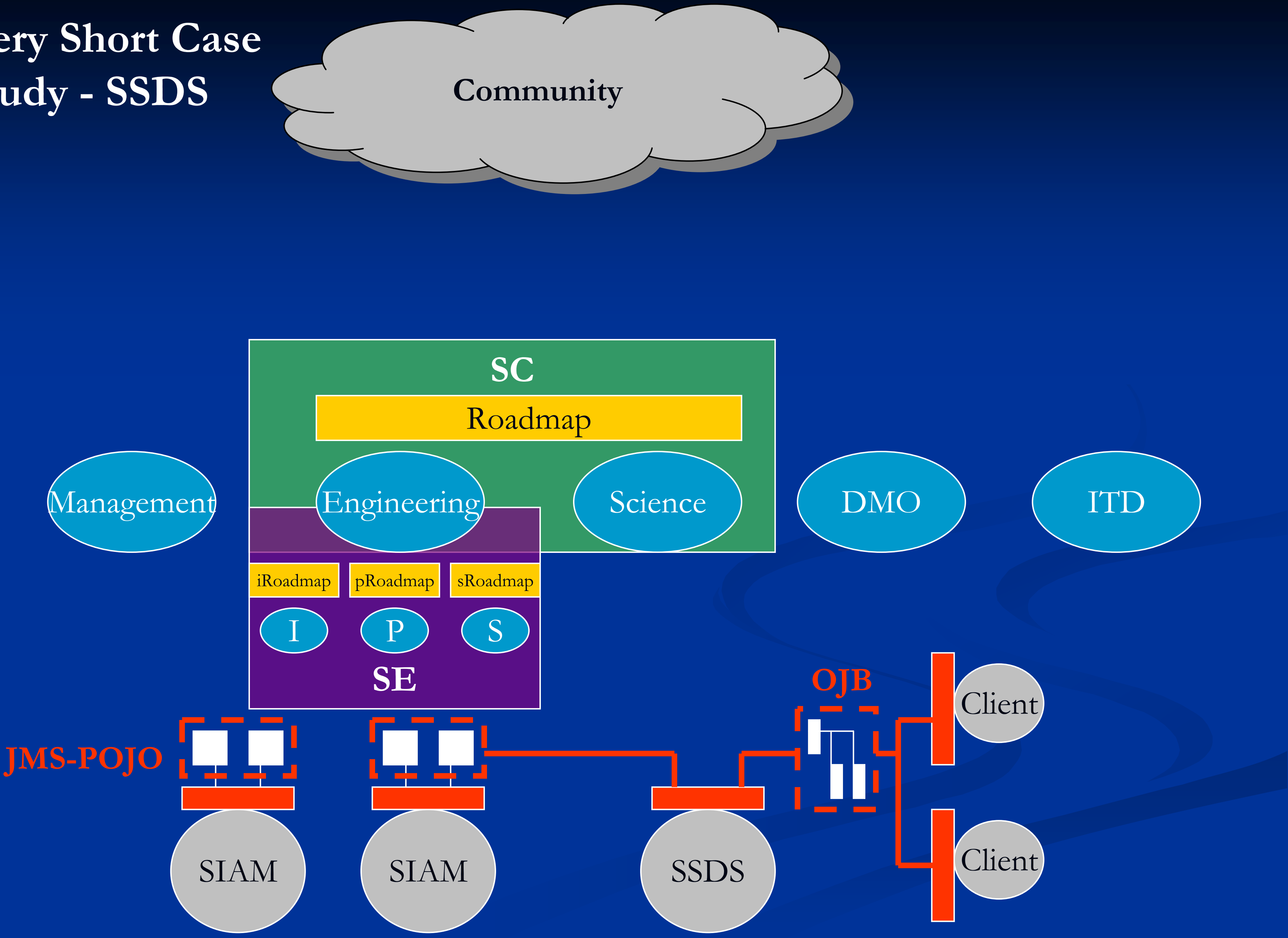
Very Short Case Study - SSDS



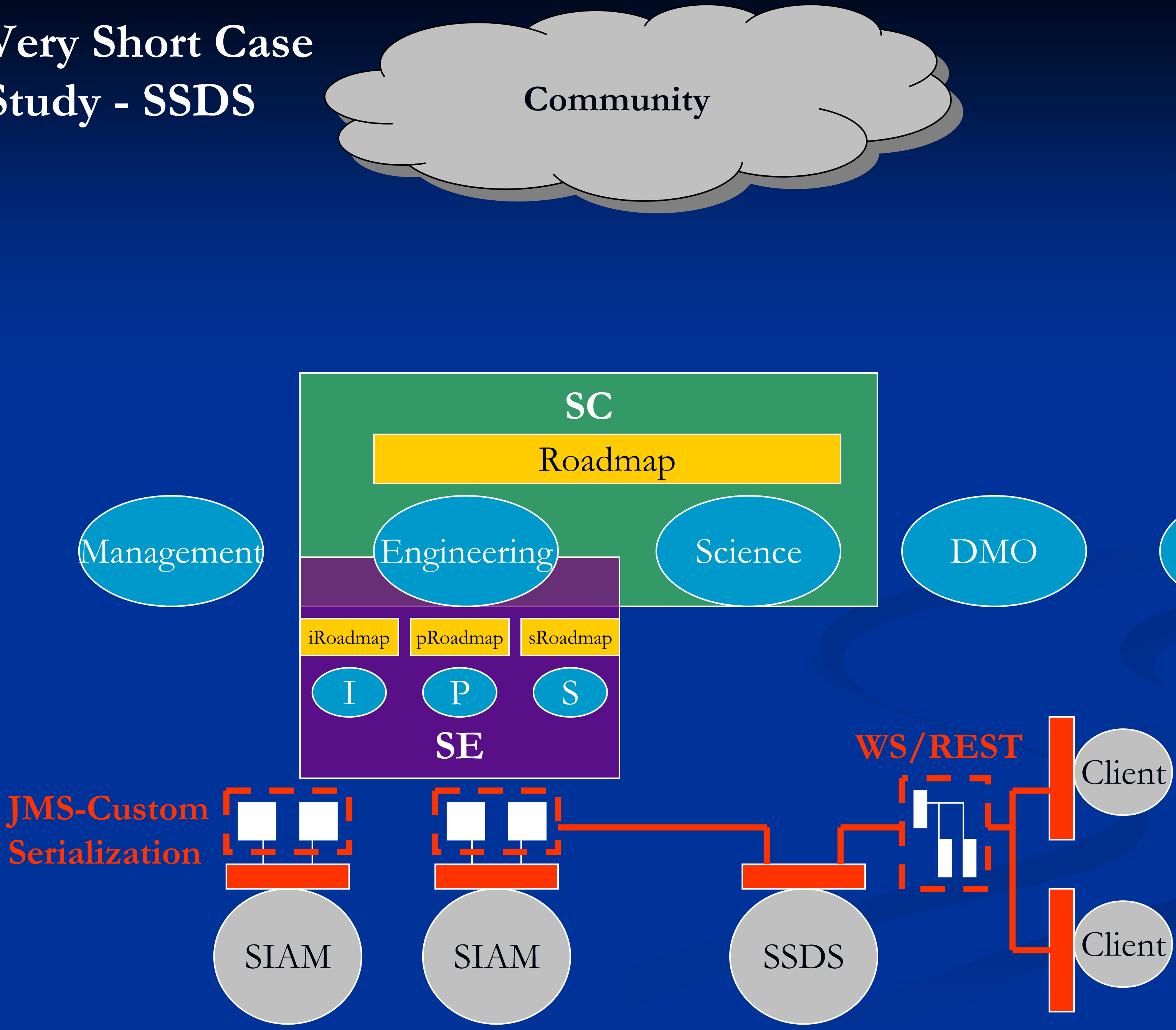
Very Short Case Study - SSDS



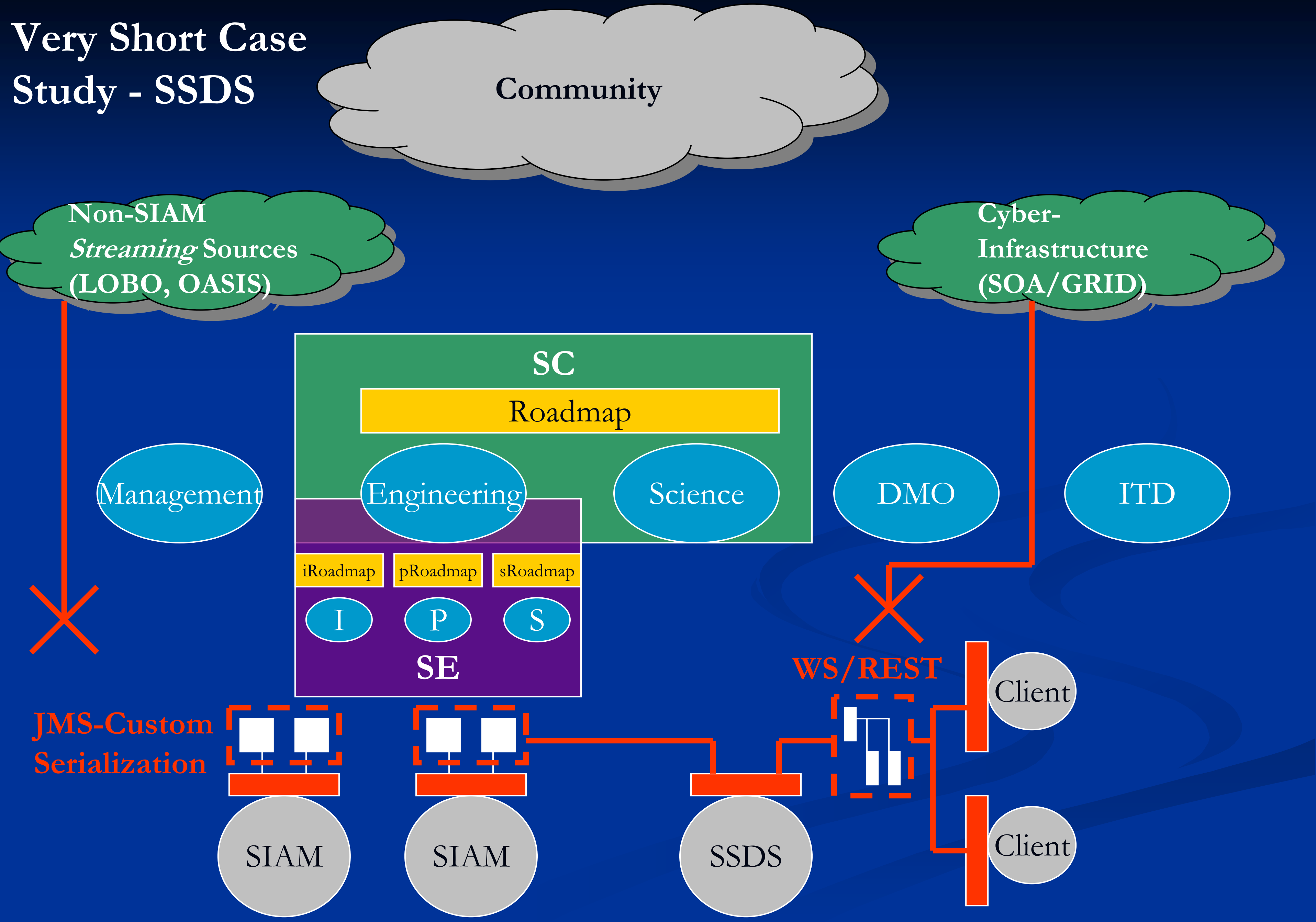
Very Short Case Study - SSDS



Very Short Case Study - SSDS



Very Short Case Study - SSDS



Next Steps

- Complete Science interviews.
- Engage Operations, Support Engineering
- Synch with group members
- Work on Strategic Engineering Council
- Work on Systems Engineering Process
- Synch with Platforms and Instruments Groups
- Distill Opportunities to concrete thrusts and capabilities
- Be ready for proposal process 2006