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

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
  - Clague, Vrijenhoek, Haddock, Paull, Robison, Scholin, Chavez, Barry
  - Discussed science themes and needs
- Scientists identified their own needs
  - Often reported to be community needs also
- Science Pull/Tech Push/Infrastructure model
  - Even scientists' needs may not be science pull
  - Goal: Move latter categories to science pull

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

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.
  - Many technical needs involved integrated projects/data
- General Strengths | Opportunities
  - Matched our own assessment (in SWOT)
  - Video, samples, benthic mapping
  - Data quantity (and context)

Community Needs: Visualization/Integration

- Way more feedback than any other topic
- Will come back to this topic after talking about the process.

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, not "Hard QC"
- Integration (of data sets) key to effective QC
- Flagging data (not deleting it) should be standard
- User participation (e.g., in flagging) would be nice
- QC of operations should be first step
  - "Got data?"

Community Needs: Visualization/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
- Visualizing a recurring theme
  - Context, multiple scales, over time
  - Make it easy to use, integrative
  - Key for context and analysis

Community Needs: Data Organization

- Wish list
  - Need easy way to "keep similar data together"
  - Really nice user interface very important
  - Quickly accessing context for data
  - Replace notebooks (must be pervasive interface)
- Déjà vu
  - Many requests already met in MBARI tools

Community Needs: Tools & Applications

- Many domain-specific requests
- Often builds on existing strengths
- Turning data into knowledge a theme

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

Approach

- Lots of advice about how we do this
- Marketing: early deliverable, make it easy
- Data Mgt: (don't) centralize data
- Strategic Planning: big picture for all
- Priorities: users vs products, new vs upgrades
- Best Practices: include users (testers), create products (with data), stay close to 'raw data'
- Data Security: Very important, make consistent
- Policies: Data should go out (unless it shouldn't)

Segue to Dan

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

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

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

Results of SWOT

- 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

Segue to Kevin

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

- True inter-group dependencies
  - A trend in Oceanography is towards integrated systems and tools that work in those integrated systems
- We risk fragmentation
  - With who?
    - 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 subgroups 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 –m more words

Management

Engineering

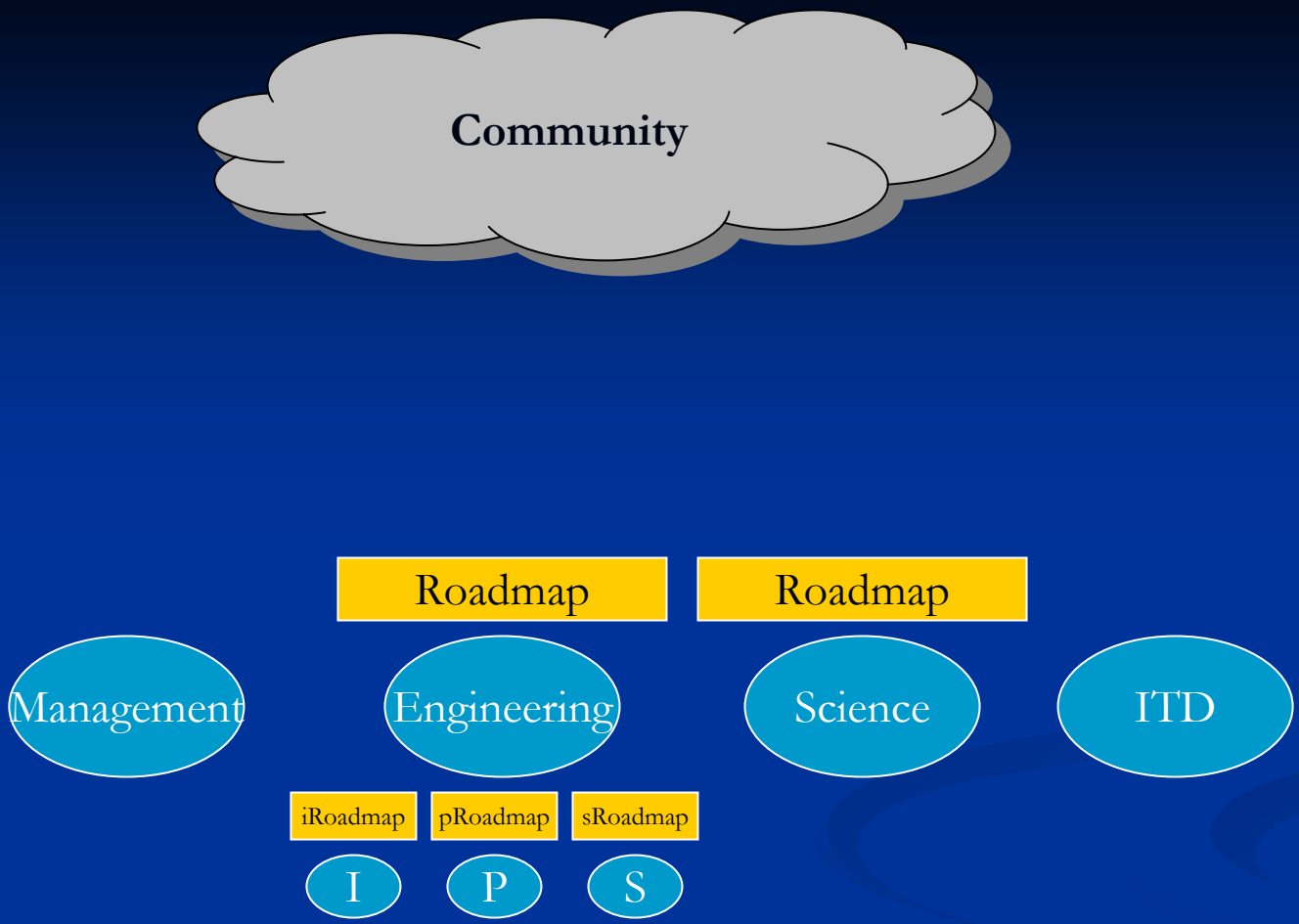
Science

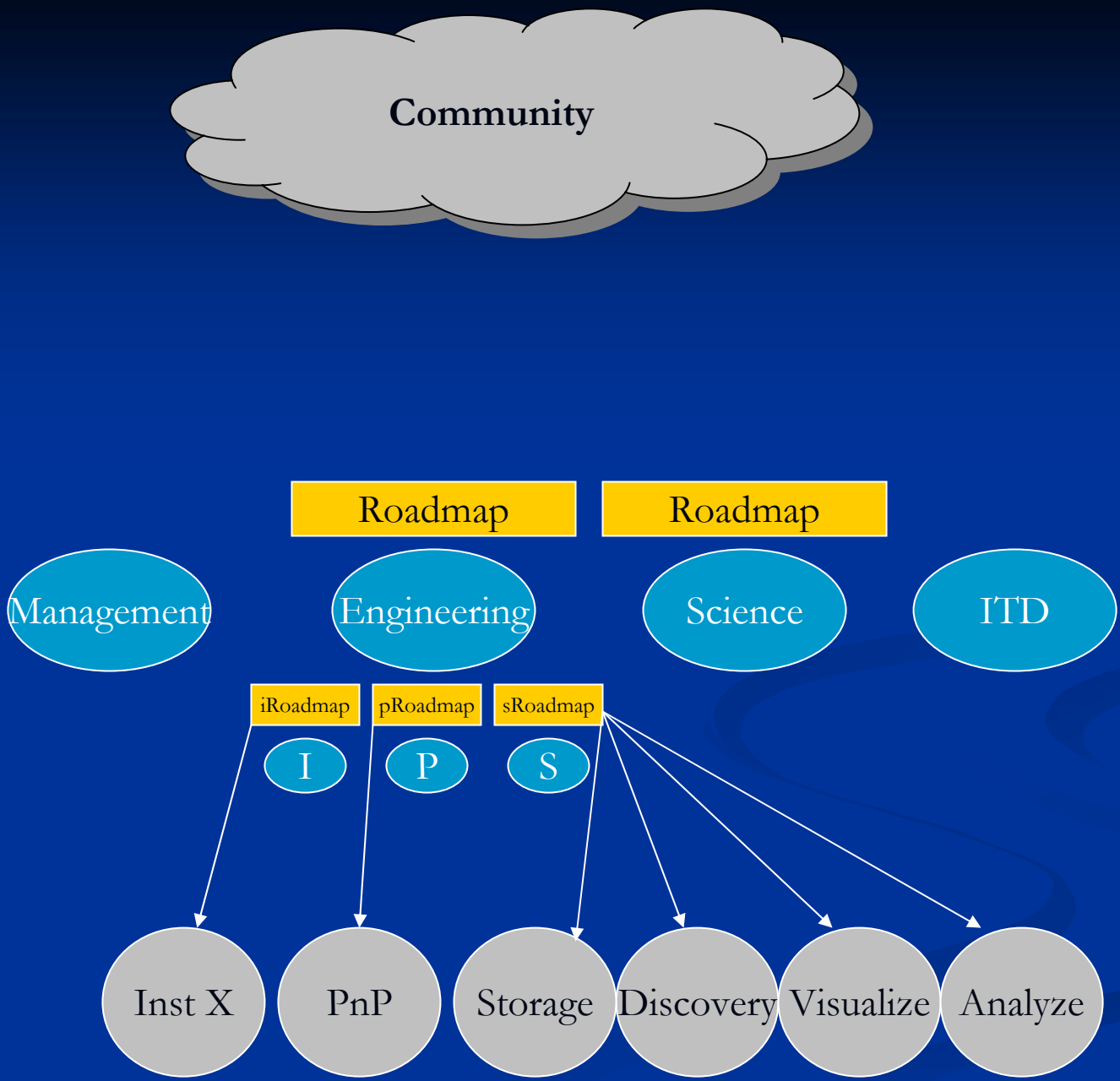
ITD

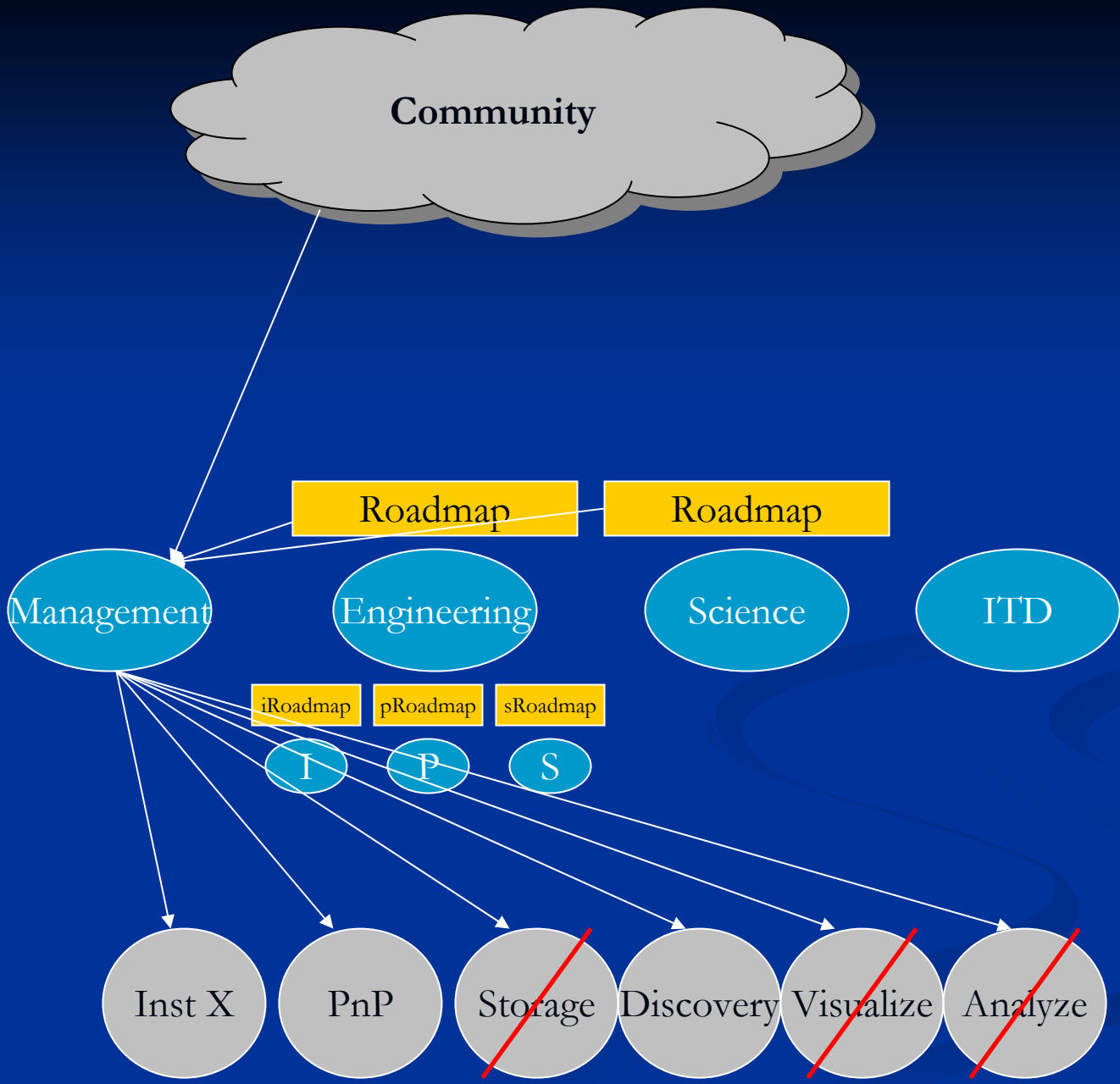
I

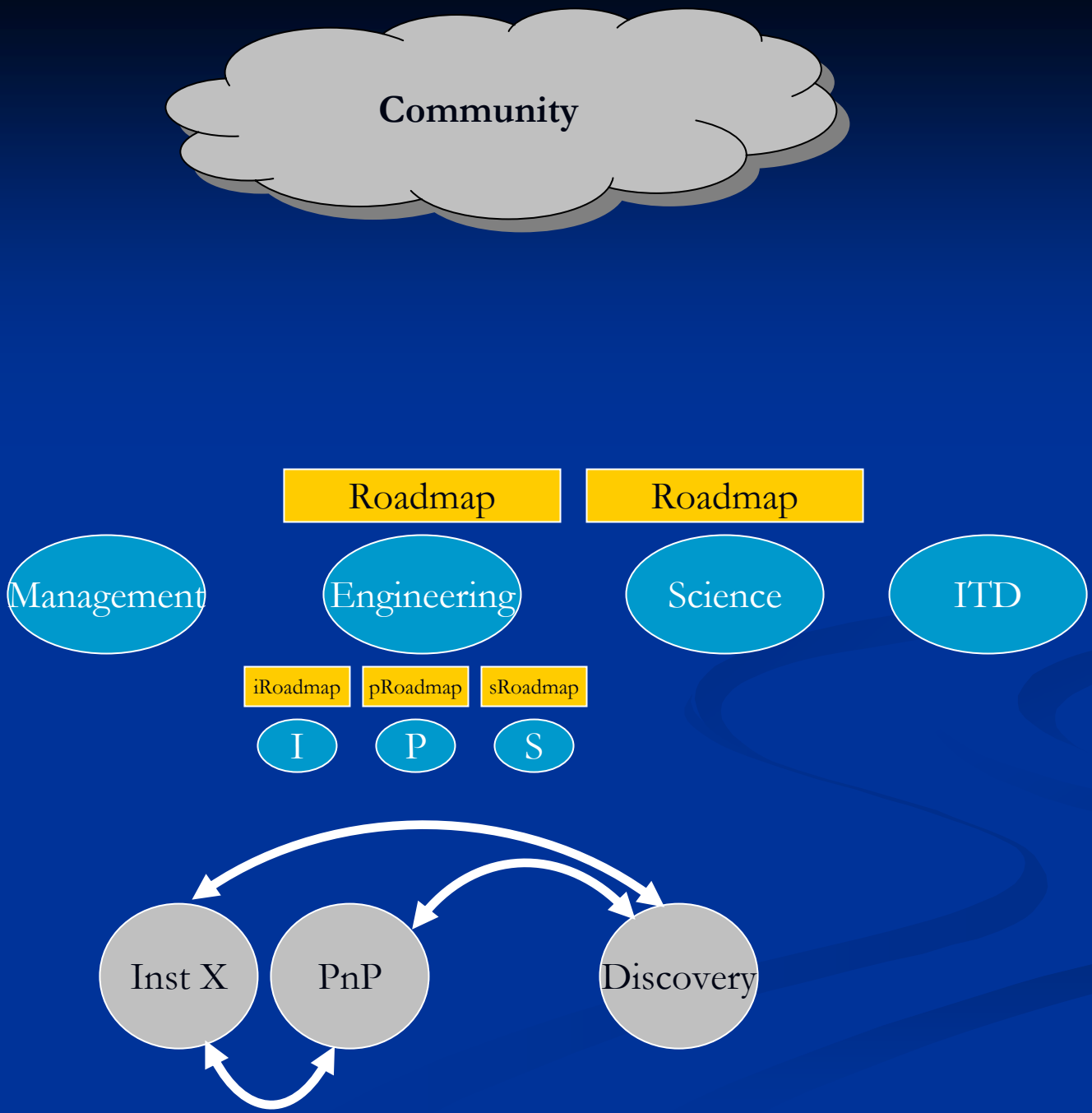
P

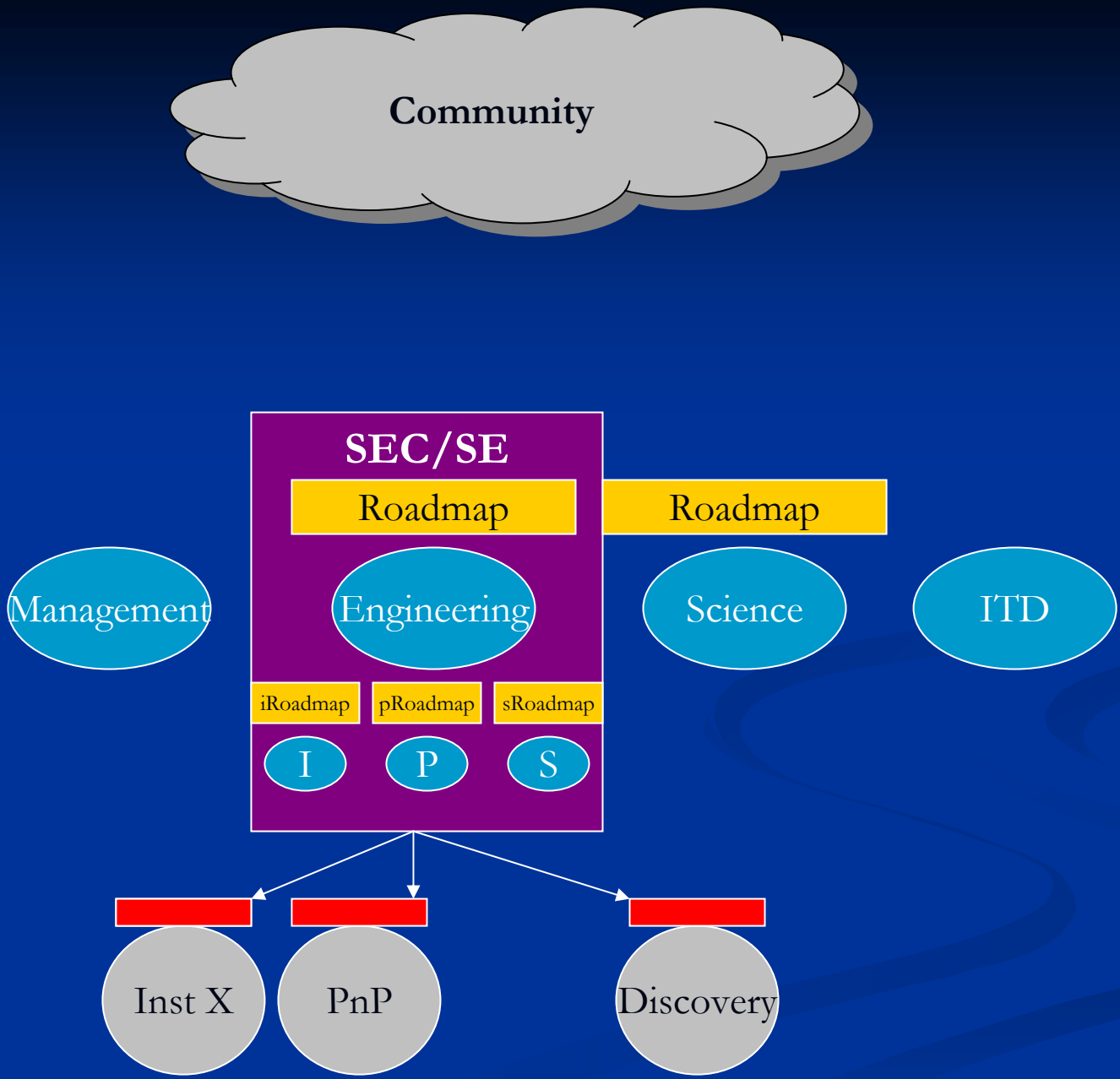
S

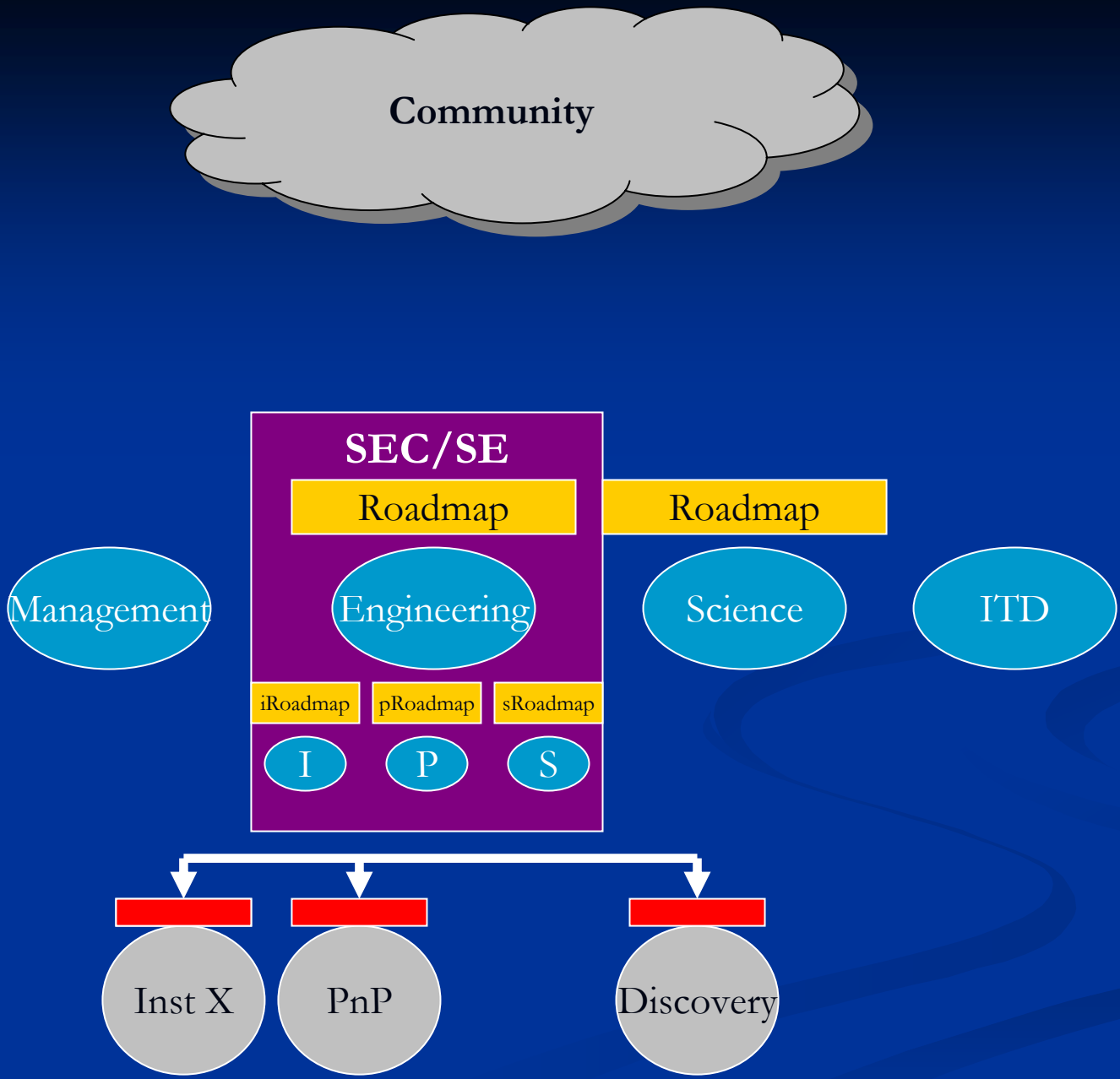

Community
Roadmap
Roadmap
Management
Engineering
Science
ITD
iRoadmap
pRoadmap
sRoadmap
I
P
S

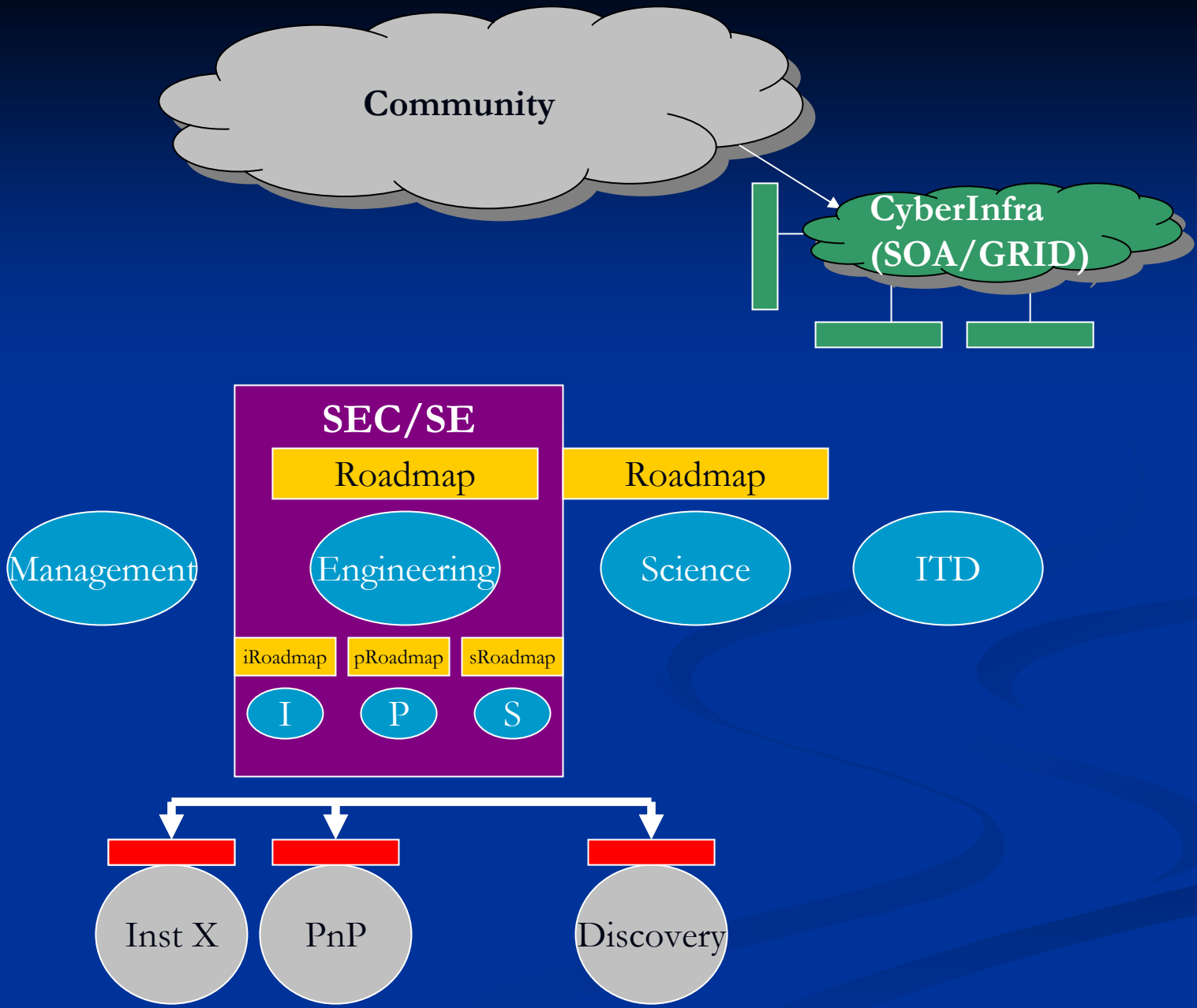

Community
Roadmap
Roadmap
Management
Engineering
Science
ITD
iRoadmap
pRoadmap
sRoadmap
I
P
S
Inst X
PnP
Storage
Discovery
Visualize
Analyze

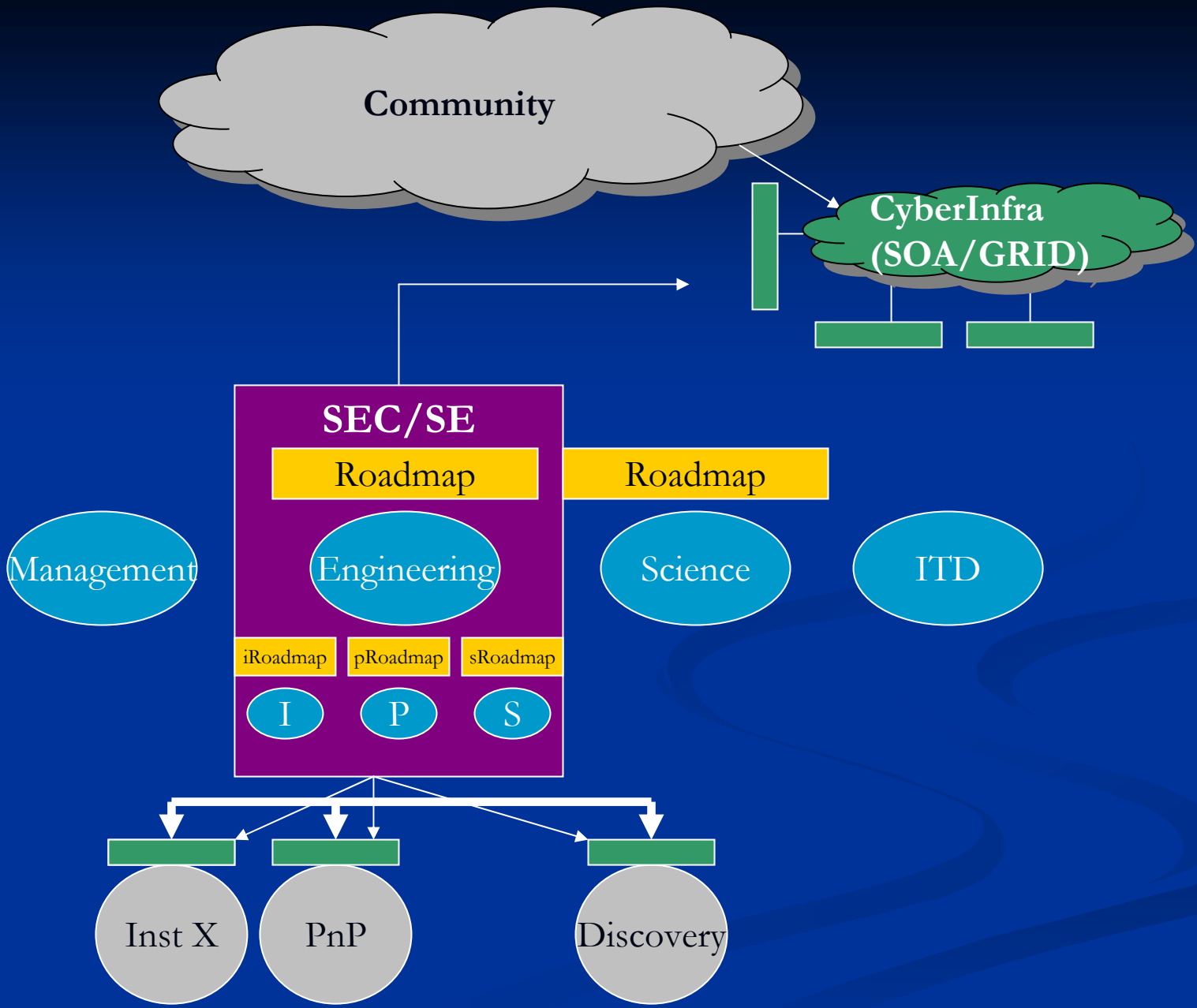

Community
Roadmap
Roadmap
Management
Engineering
Science
ITD
iRoadmap
pRoadmap
sRoadmap
I
P
S
Inst X
PnP
Storage
Discovery
Visualize
Analyze

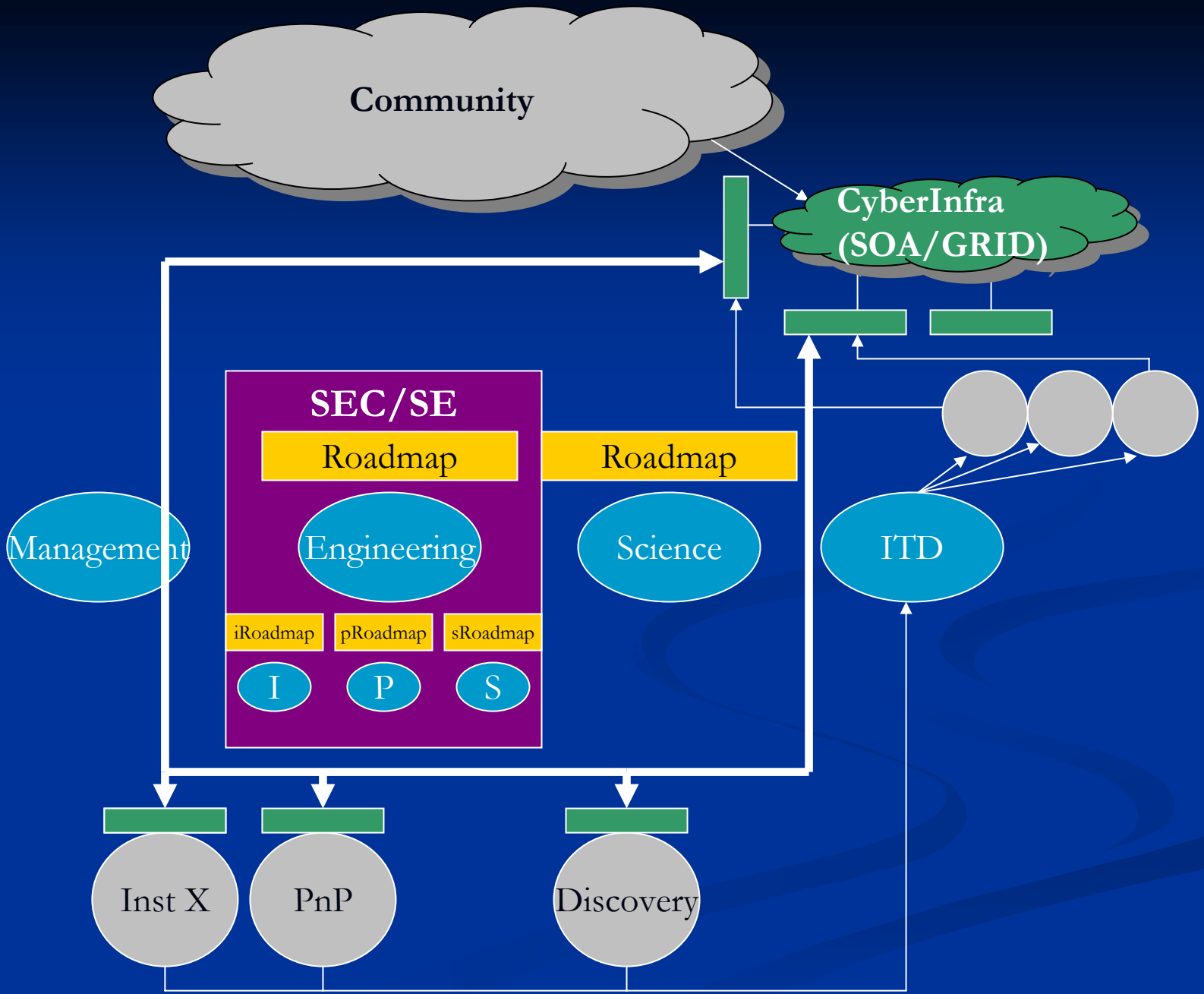

Community
Roadmap
Roadmap
Management
Engineering
Science
ITD
iRoadmap
pRoadmap
sRoadmap
I
P
S
Inst X
PnP
Discovery


Community
SEC/SE
Roadmap
Roadmap
Management
Engineering
Science
ITD
iRoadmap
pRoadmap
sRoadmap
I
P
S
Inst X
PnP
Discovery

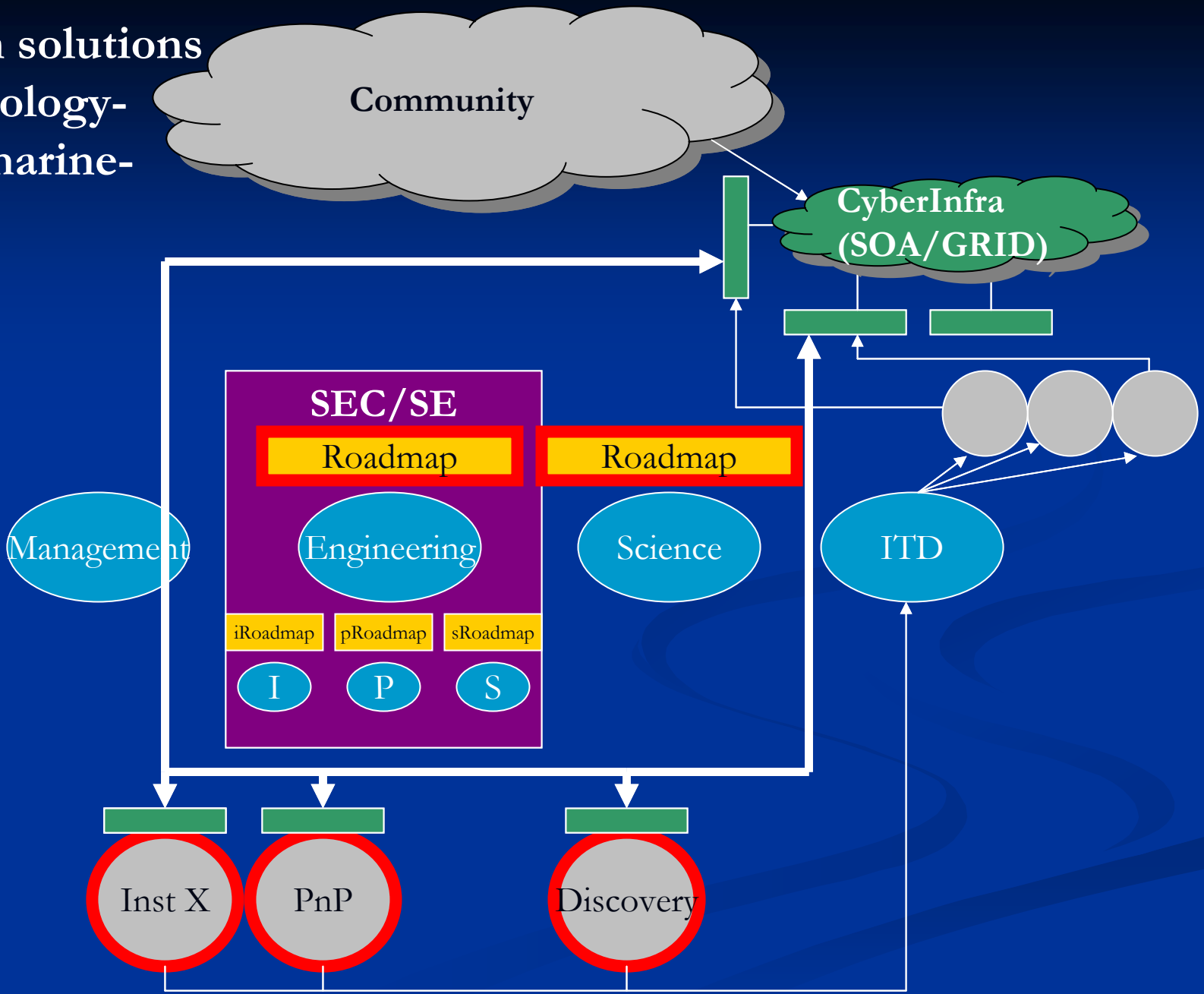

Community
SEC/SE
Roadmap
Roadmap
Management
Engineering
Science
ITD
iRoadmap
pRoadmap
sRoadmap
I
P
S
Inst X
PnP
Discovery


Community
CyberInfra
(SOA/GRID)
SEC/SE
Roadmap
Roadmap
Management
Engineering
Science
ITD
iRoadmap
pRoadmap
sRoadmap
I
P
S
Inst X
PnP
Discovery

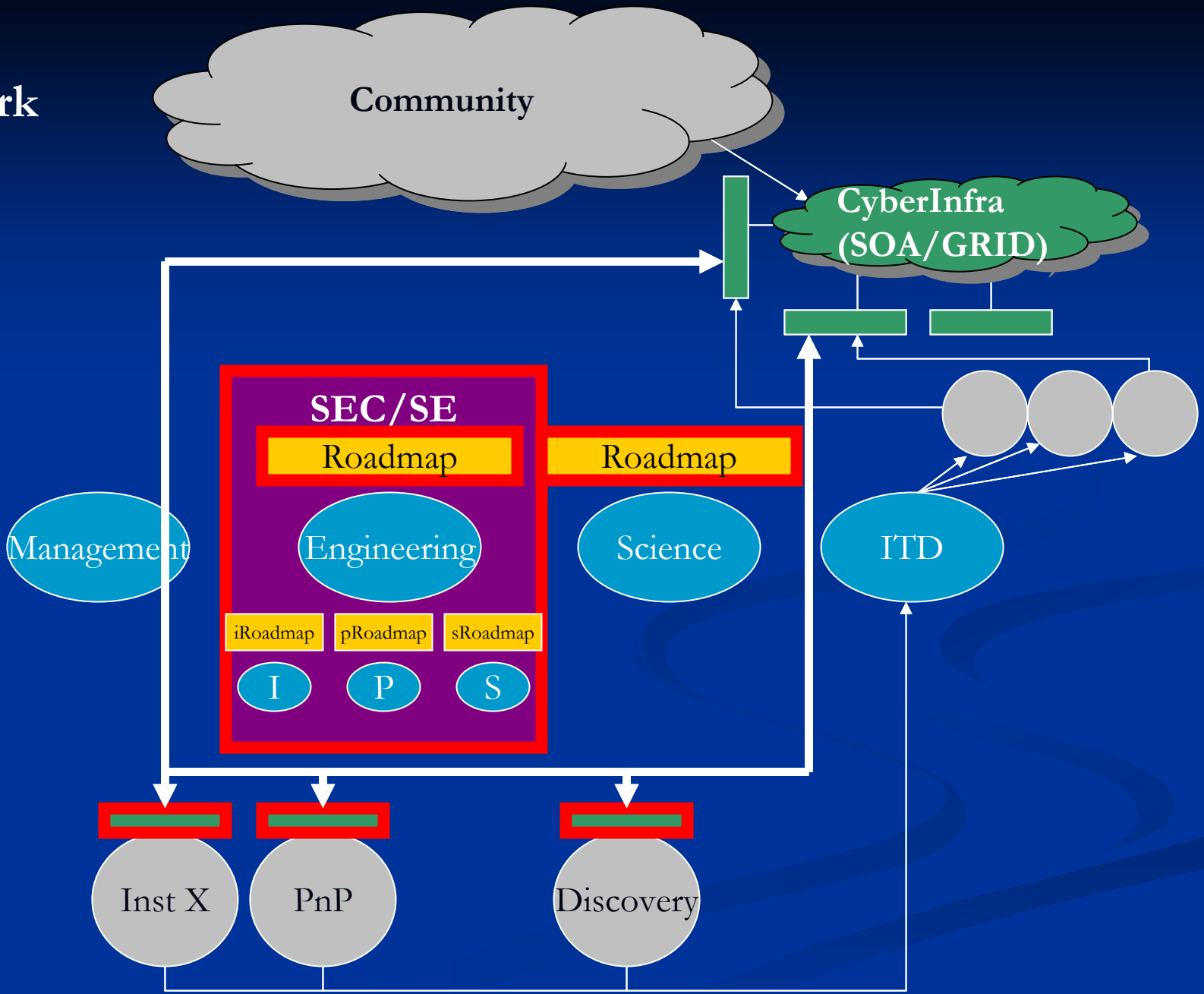

Community
CyberInfra
(SOA/GRID)
SEC/SE
Roadmap
Roadmap
Management
Engineering
Science
ITD
iRoadmap
pRoadmap
sRoadmap
I
P
S
Inst X
PnP
Discovery


Community
CyberInfra
(SOA/GRID)
SEC/SE
Roadmap
Roadmap
Management
Engineering
Science
ITD
iRoadmap
pRoadmap
sRoadmap
I
P
S
Inst X
PnP
Discovery

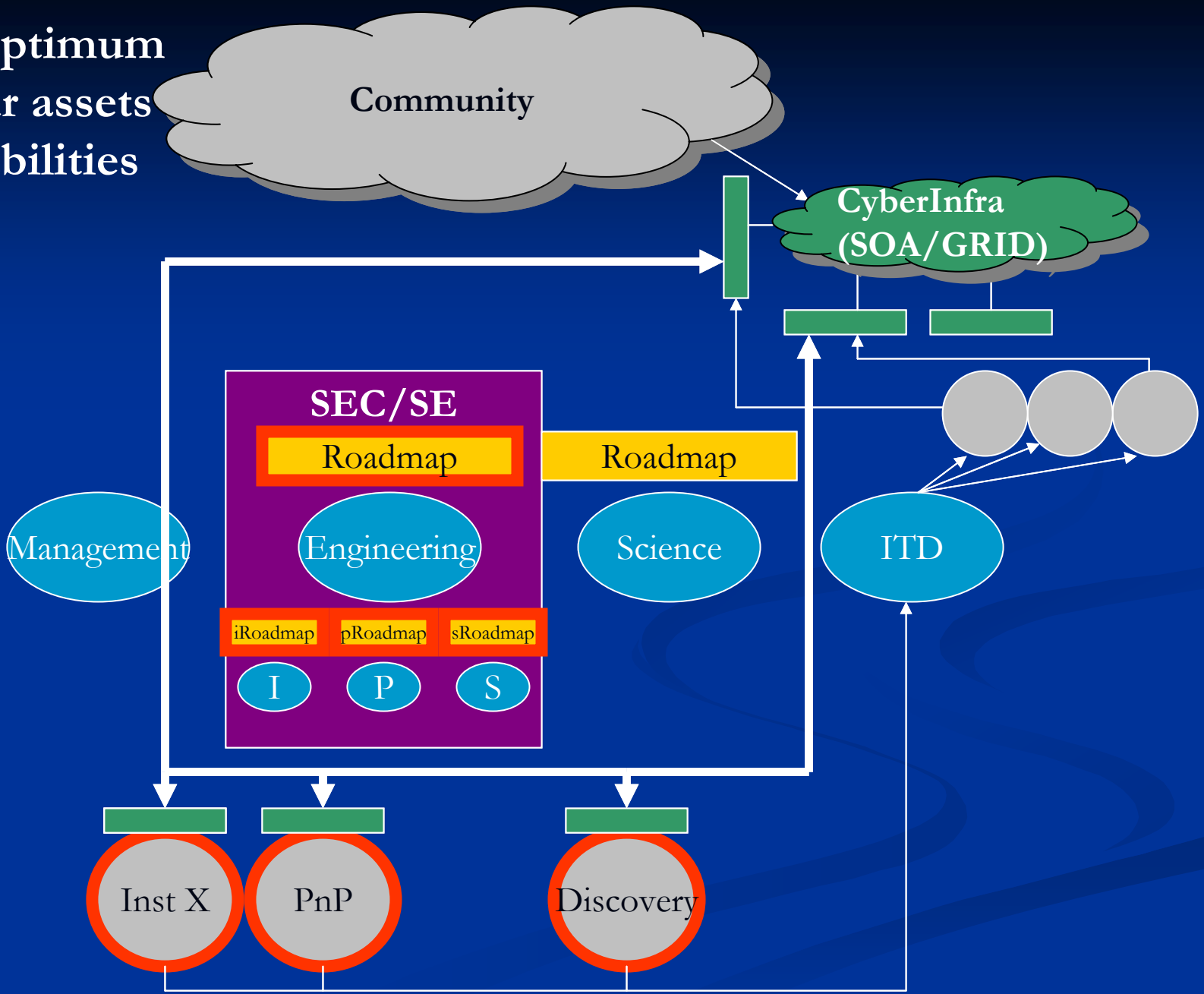
Focus on solutions for technology-limited marine-science



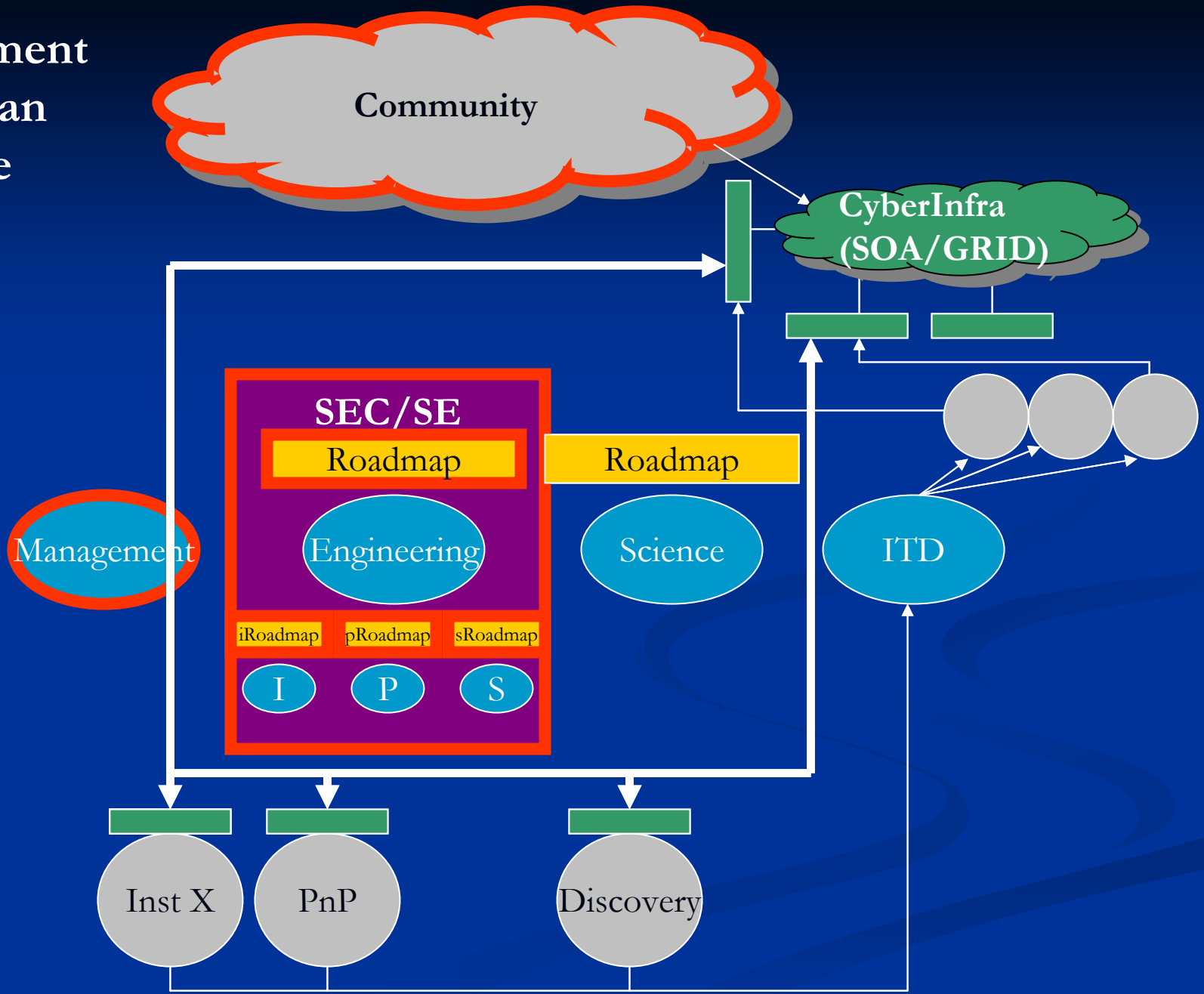
Insist on Teamwork



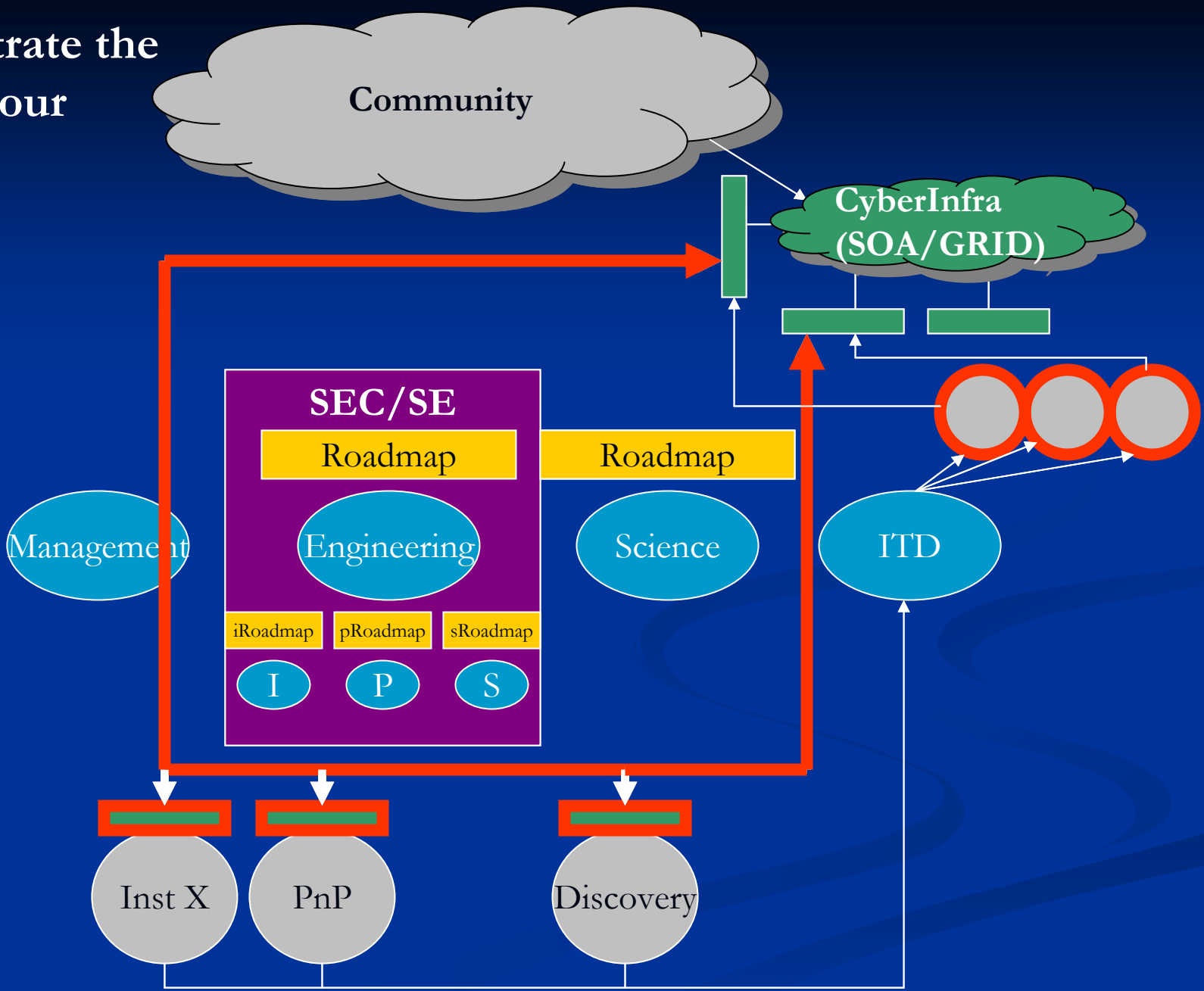
Ensure optimum use of our assets and capabilities



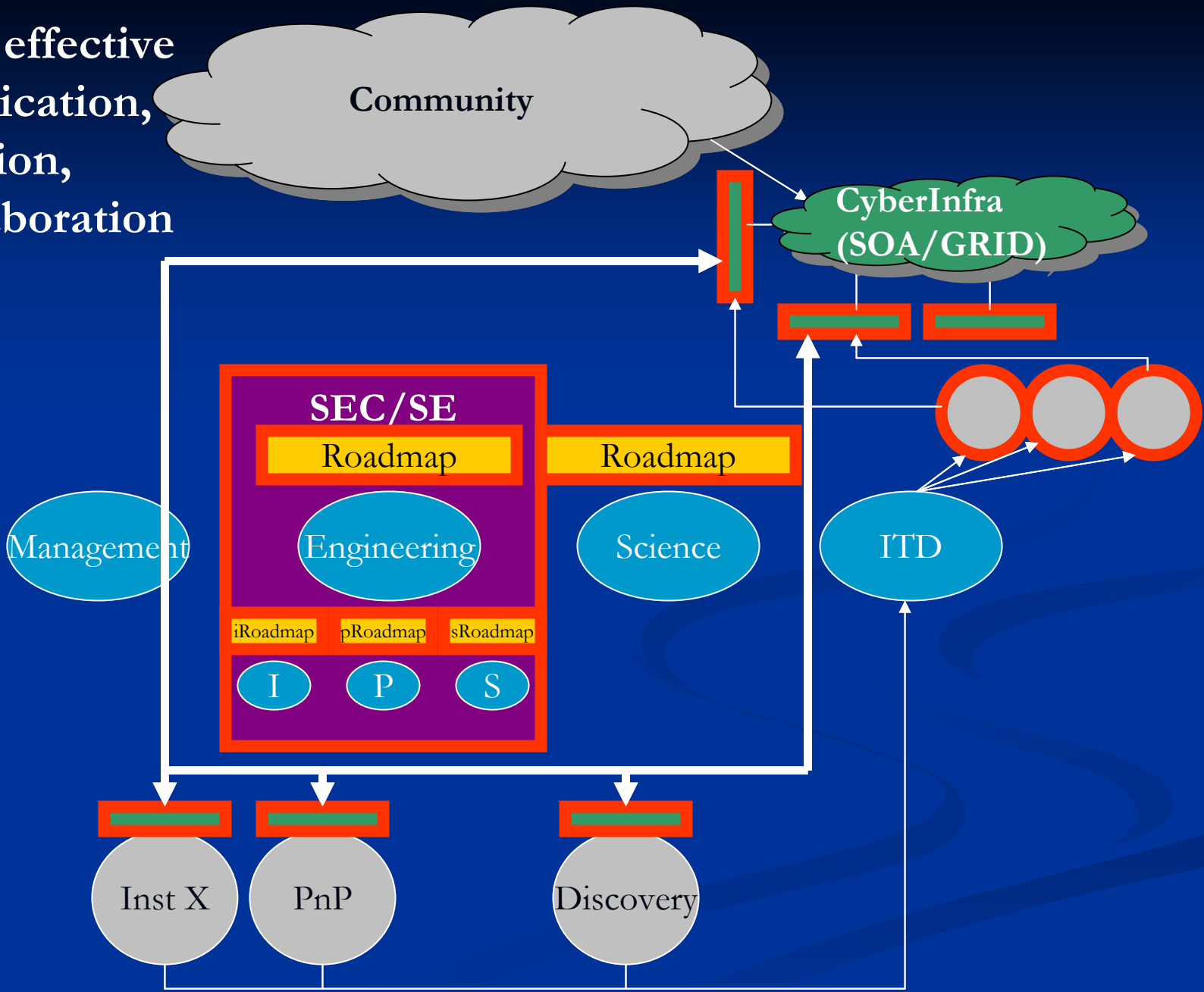
Complement rather than duplicate



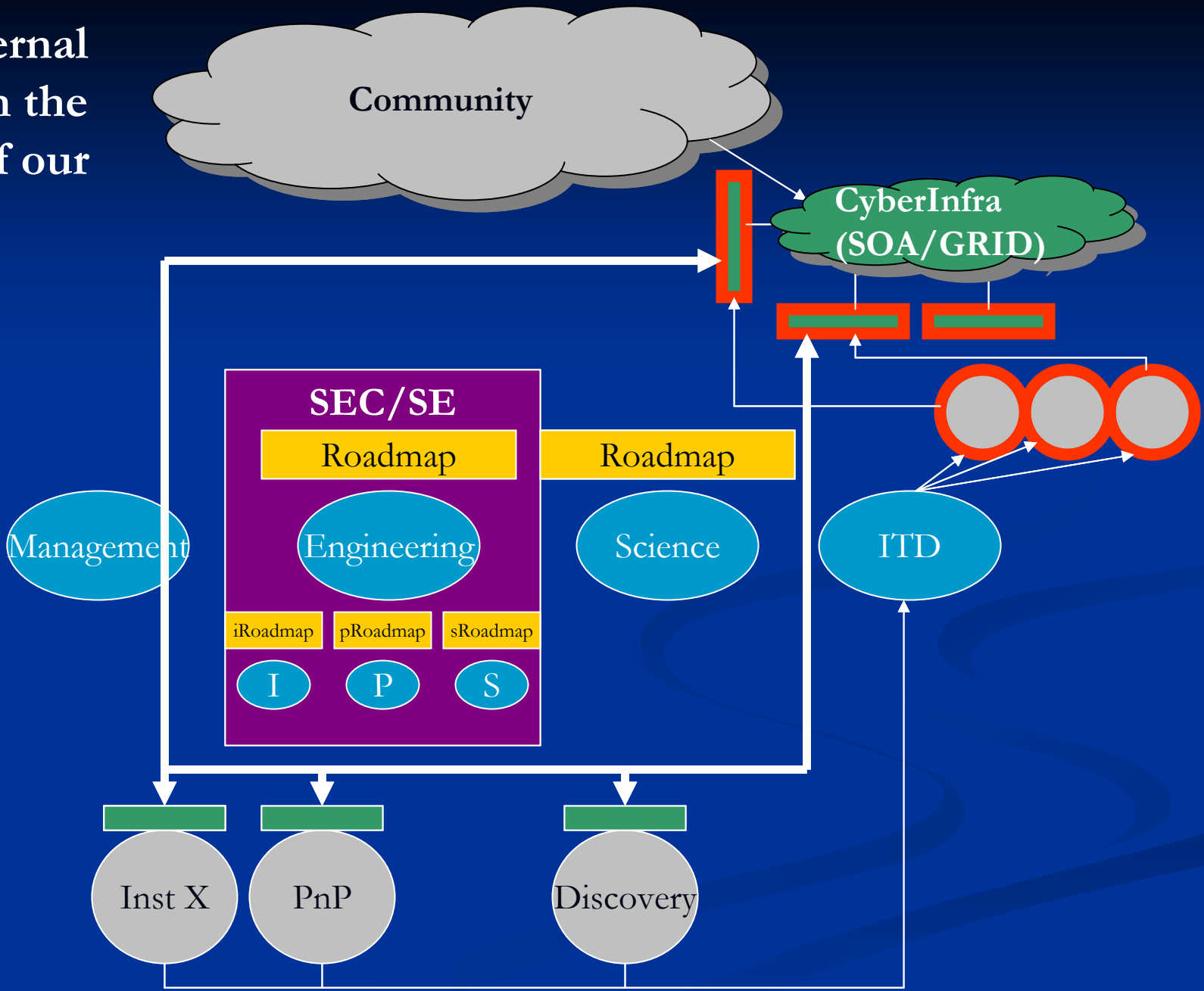
Demonstrate the worth of our efforts



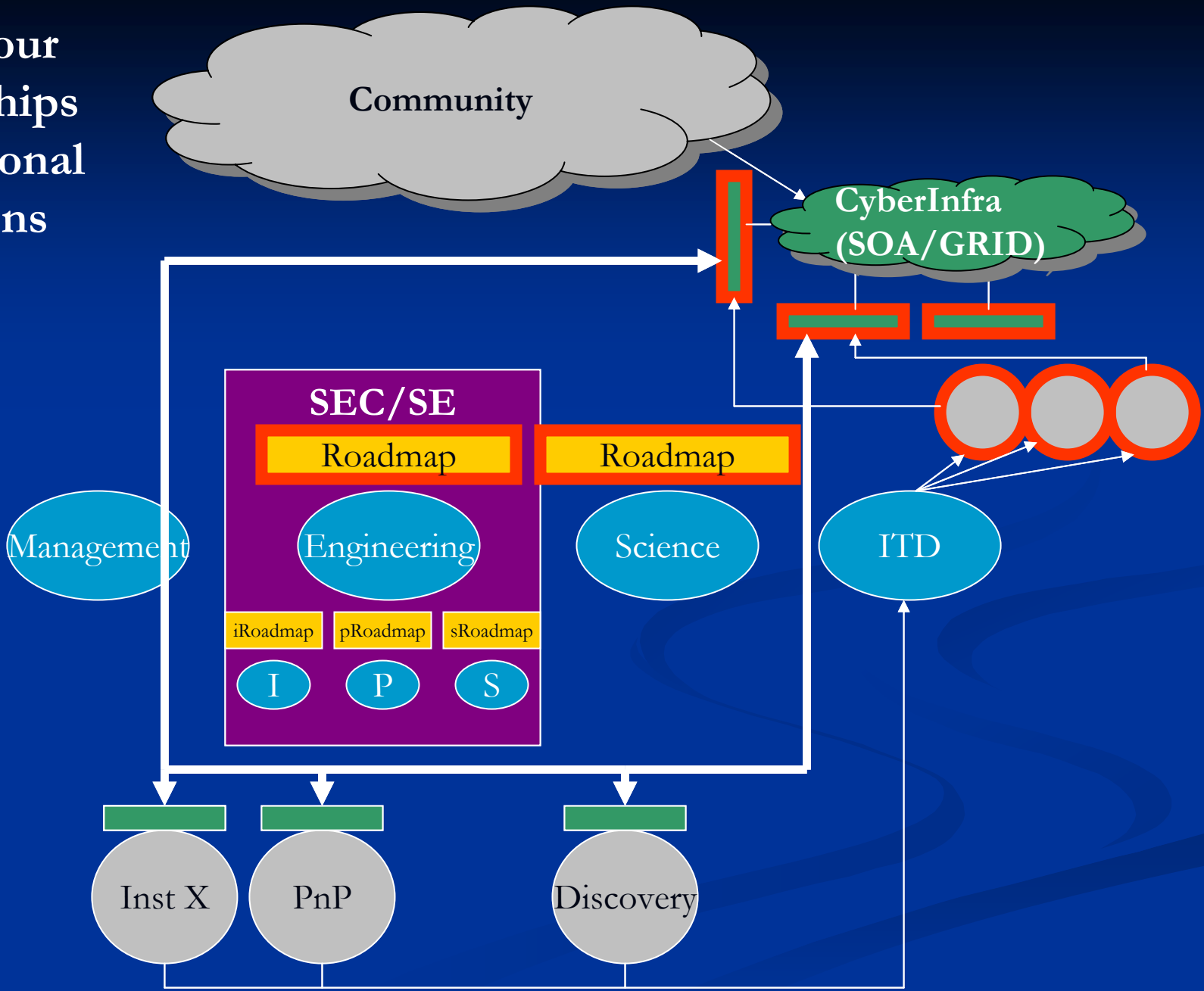
Promote effective communication, cooperation, and collaboration



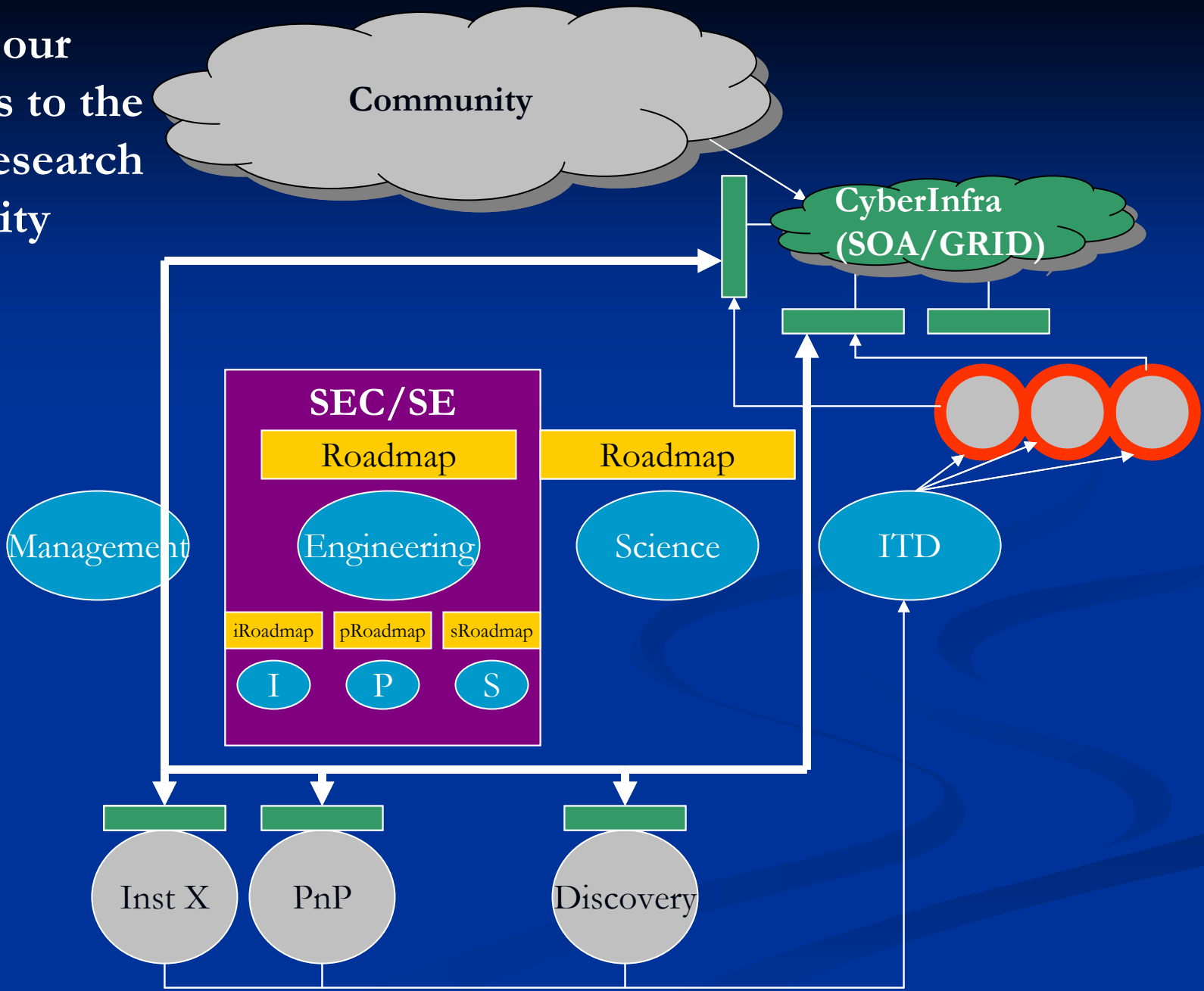
Seek external advice on the quality of our work



Expand our relationships with regional institutions



Transfer our successes to the marine research community



Communicate with the public