

**Scenario Approach to Question-
Answering**

ARDA – NRRC Workshop

June 23 – 27, 2003

Elizabeth D. Liddy

**Center for Natural Language Processing
Syracuse University**

December 5, 2003

Workshop Goals

1. Increase the QA research community's understanding of the **information-seeking needs & cognitive processes of Intel Analysts.**
2. Determine the **question-answering technology capabilities** that Analysts' scenario-based information-seeking tasks require.
3. Develop a **research agenda** for advancing current state-of-the-art QA to support scenario-based intelligence analysis.

Attendees

- **11 Intel Analysts**
 - **4 National Security Agency (NSA)**
 - **4 Defense Intelligence Agency (DIA)**
 - **2 Center for Non-Proliferation Studies (CNS)**
 - **1 US Navy**
- **16 Question-Answer Researchers**
 - **10 university (6 Faculty; 4 PhD Students)**
 - **CMU, Iowa, Michigan, MIT, Rutgers, Syracuse, USC**
 - **6 corporate**
 - **Boeing, CL Research, IBM, Mitre, Tarragon**

Data & Systems

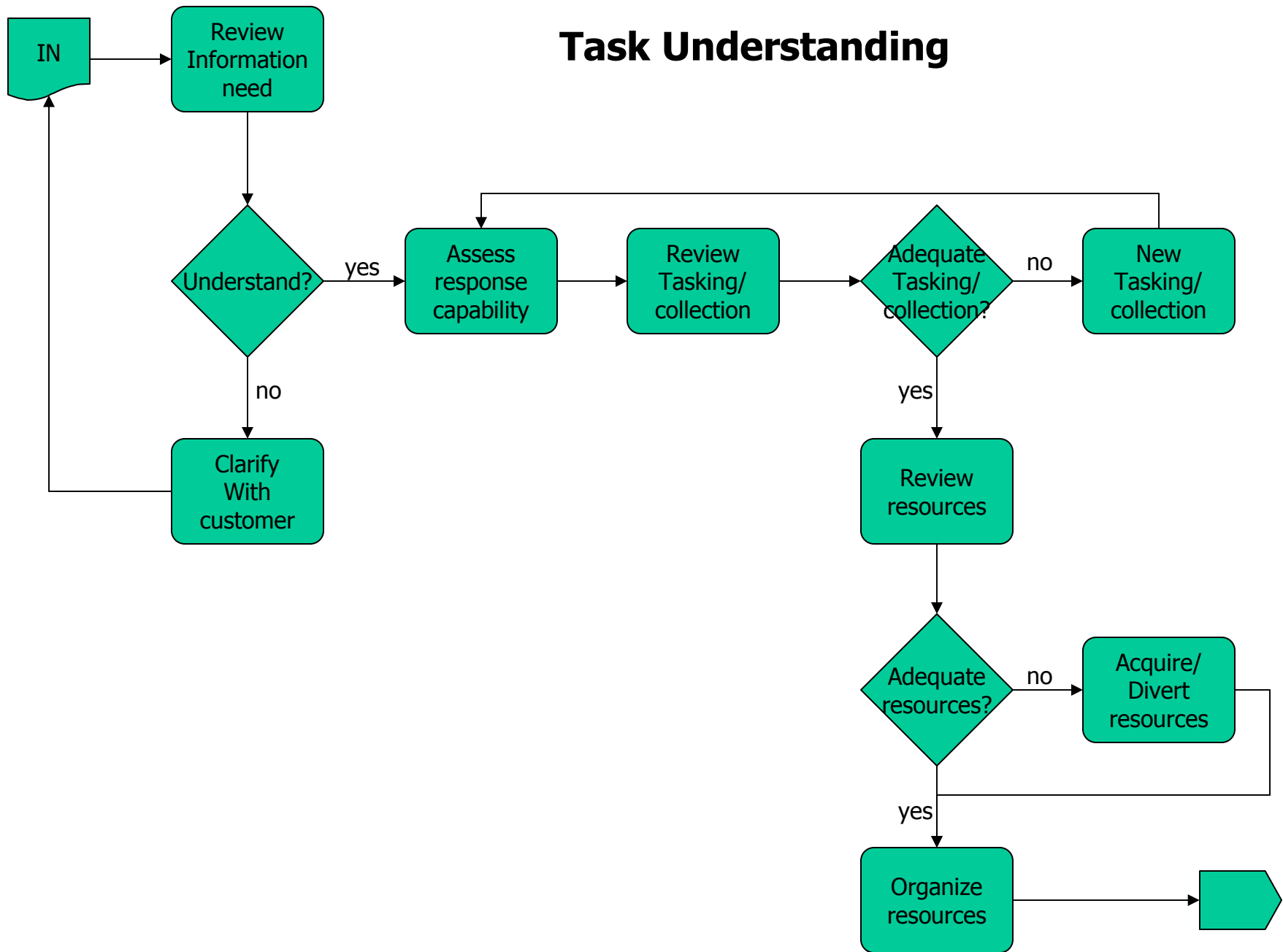
- **Data collection provided by Center for Non-Proliferation Studies**
- **CNS-developed scenarios / questions based on realistic interests of analysts**
 - **4 full-fledged Scenarios**
 - **12 shorter Questions**
- **4 QA systems – IBM, ISI, LCC, Mitre**

	Phase 1	Phase 2	Phase 3
Goal	To understand Intel Analysts' processes in scenario-based tasks.		
Task	Elicit and observe analysts' cognitive & physical actions in accomplishing scenario task via think-aloud protocols and annotations.		
Deliverable	Models of actions and processes that analysts perform.		

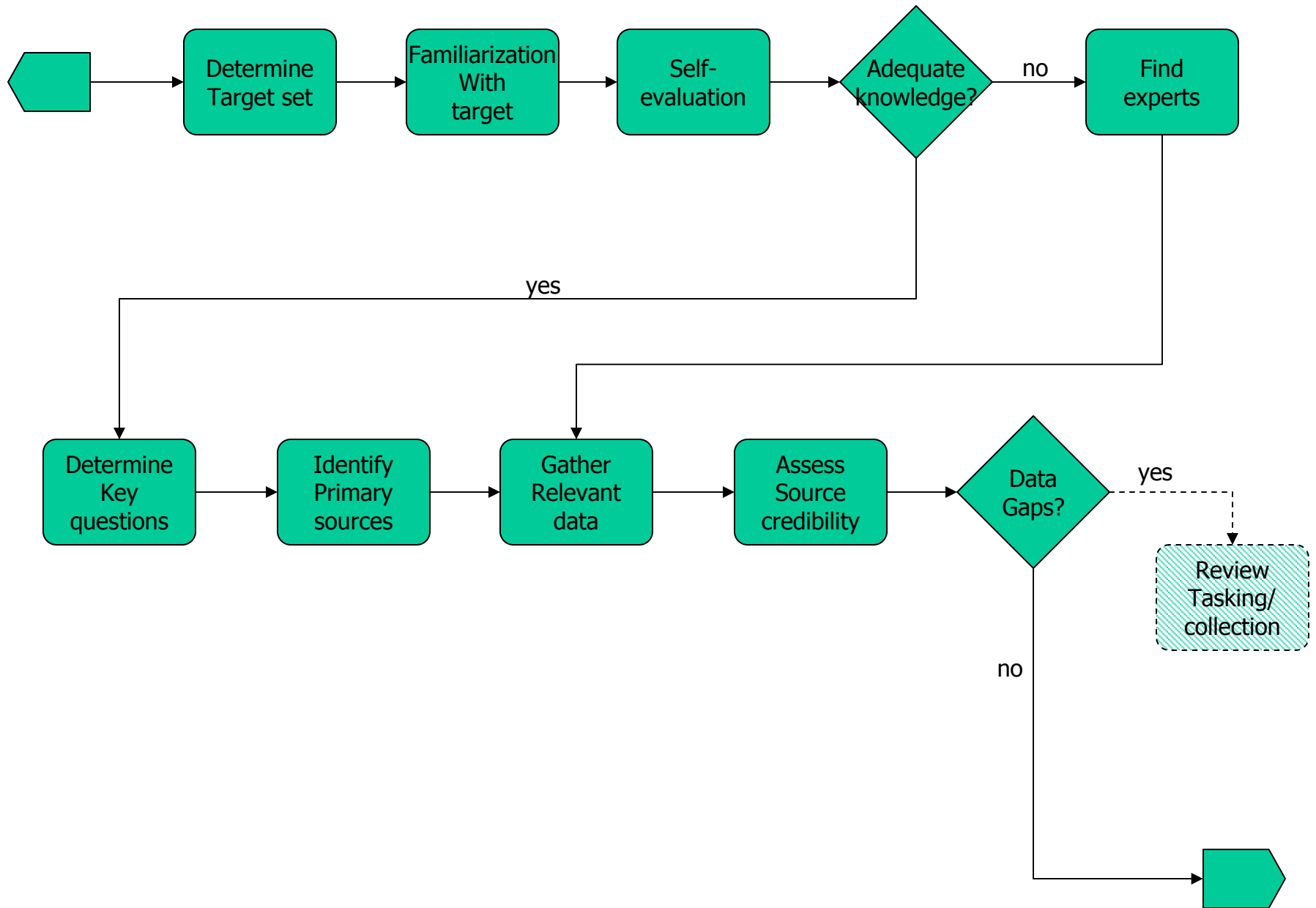
Phase I Process

- **3 teams worked separately & then reported out for group discussion**
- **Given a scenario, analysts described in detail how they would proceed**
- **Researchers took notes and prodded analysts for more information**
 - **Many pages of annotations of ‘think aloud’ protocols**
- **Basis for model & recommendations**

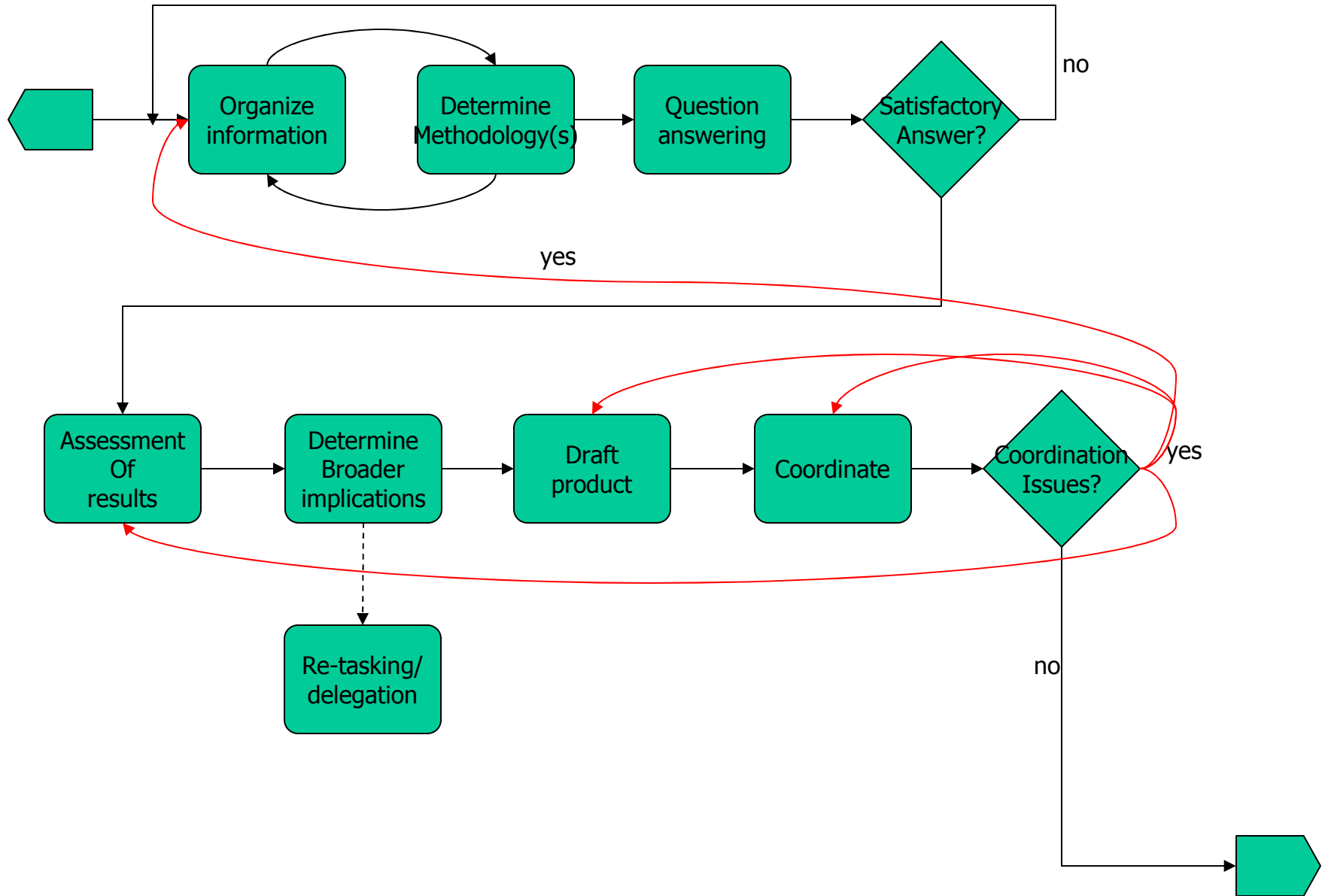
Task Understanding



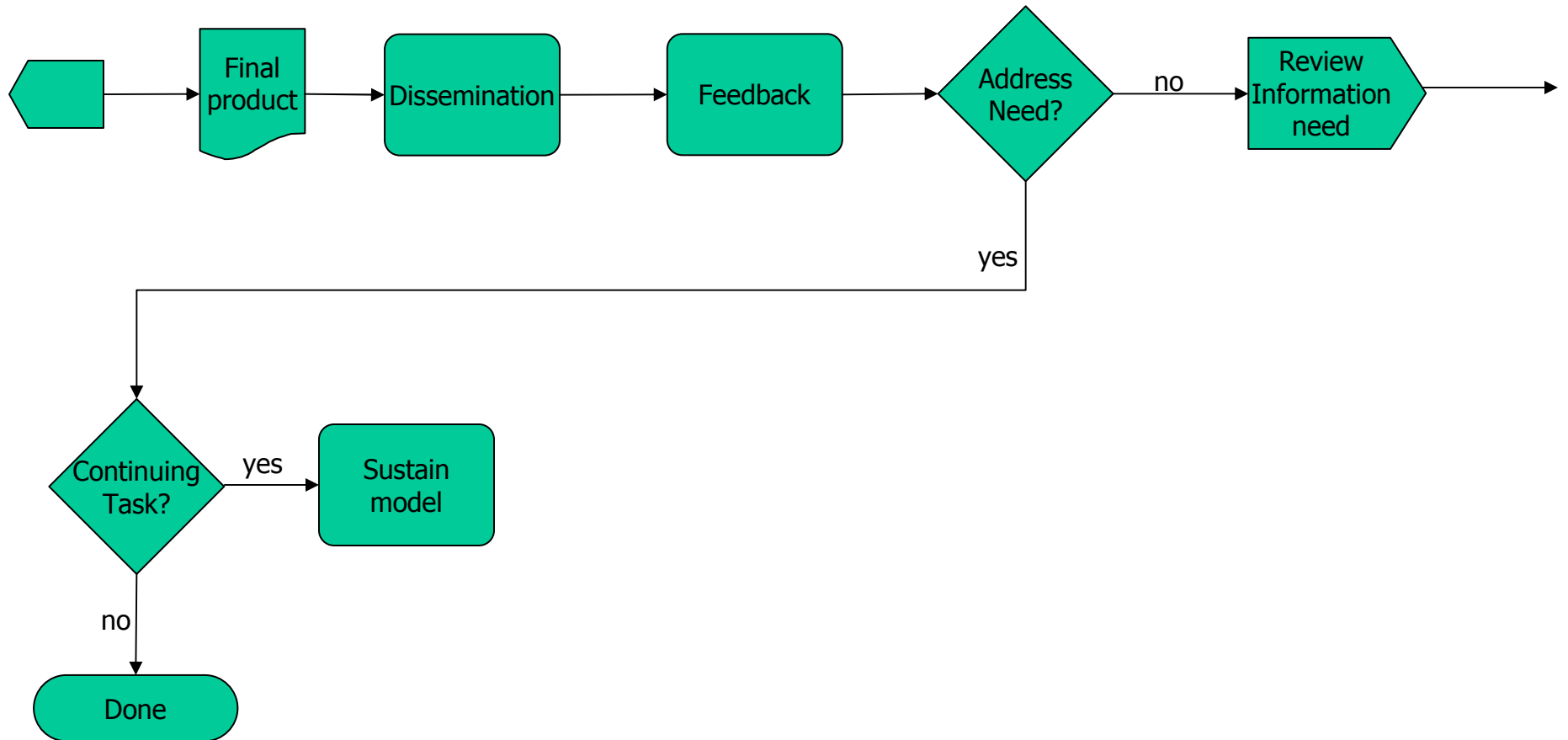
Discovery



Analysis and Synthesis



Presentation





	Phase 1	Phase 2	Phase 3
Goal	To understand Intel Analysts' processes in scenario-based tasks.	To observe the QA capabilities that analysts want to see in systems.	
Task	Elicited / observed analysts' cognitive & physical actions in accomplishing scenario task via think-aloud protocols.	Analysts used 4 QA systems & researchers interacted about how / what technologies might better assist them.	
Deliverable	Models of actions and processes that analysts perform.	Analysts' requirements / desiderata for QA technologies.	

Phase II Process

- **Teams mainly used questions generated in Phase I**
 - **Asked them of 1 or more of the 4 QA systems**
 - **Verbalized what they wanted a QA system to do for them for that task**
- **Group report-outs / discussion of QA task added to desired capabilities**
 - **Somewhat based on capabilities missing from / suggested by QA systems used**
 - **More based on what analysts would find useful to their processes**

Analyst Desiderata Ranking Process

Used a modified Delphi process:

- 1. At workshop**
 - 1. 40 desiderata were produced after 4 days**
 - 2. Analysts ranked their top 10**
 - 3. These were presented at mid-term review**
- 2. After workshop, desiderata were reviewed, merged, elucidated by a sub-group**
- 3. Analysts ranked their top 20 of the revised set of 30 desiderata**

Analysts' Top 20 Requirements

1. **Templates** of commonly needed information for scenarios
 - Analysts can launch template as a query for system to search and fill in slots along with links back to sources
 - Sample templates are Biography, Order of Battle, Travel, Social Network
2. System identifies and utilizes **relations** occurring in text, permitting the analyst to query by relations, not just topic or entity
(e.g. I am looking for persons [X] who traveled from country Y to country Z.)
3. Answer will reveal the system's **line of reasoning** and justification for scoring and ranking of responses to question
4. A **single system** that eliminates need for multiple interfaces for asking the same question of multiple systems

Analysts' Top 20 Requirements (cont'd)

- 5. System indicates where information came from –
provenance**
- 6. System dialogs with analyst to clarify question**
 - enabling analysts to ask quite natural questions, while
 - ensuring that system will clarify question when not clear, as would a reference librarian.
- 7. In response to a question, system locates individual
experts and expert communities using an inter-agency
database of who is / has worked on question.**
- 8. Conceptual querying**
 - System represents analyst's need without the analyst having to adjust their query to how the system wants it.

Analysts' Top 20 Requirements (cont'd)

9. System understands all **temporal references** (e.g. *12/8/95, yesterday, current, as of...*) ;
 - time-stamps the information
 - timelines it
 - can then indicate what is the **most current** information
10. **Summarization** of results from multiple sources, with ability to drill down into results that were summarized.
11. Indication of the **reliability** / believability of a source based on analysts' past experience with the source
12. System deals with analyst's concern for strategic surprises by recognizing **unusual patterns of activities and relationships** and alerting analyst.

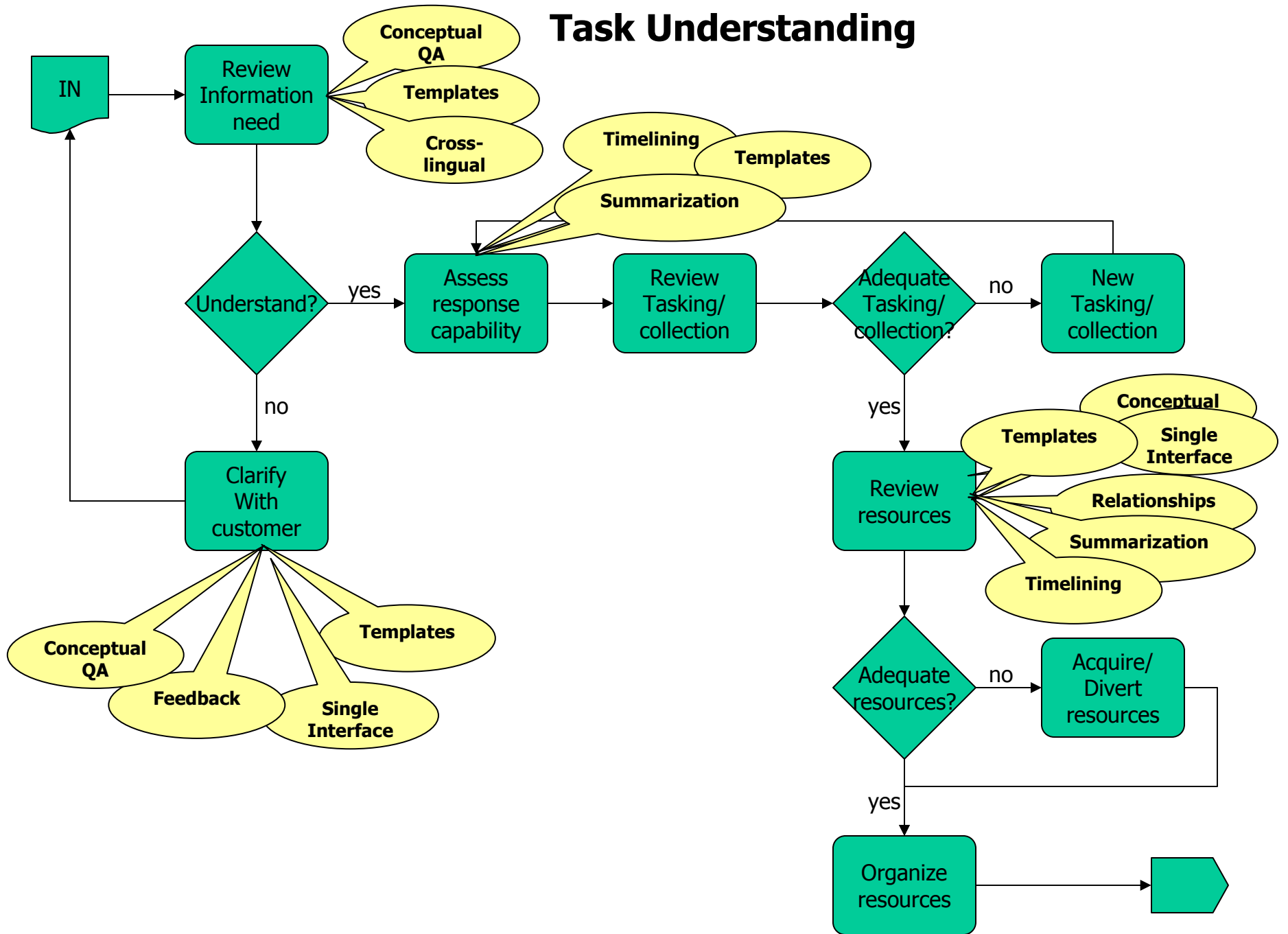
Analysts' Top 20 Requirements (cont'd)

13. **Cross-lingual** question-answering capability.
14. System recognizes **conceptual / logical equivalents** of responses
 - indicates why responses are deemed equivalent
 - if different, system highlights what the responses agree on, what they don't agree on, and what is omitted.
15. Questions of sustained interest continue to fire automatically in background as **“standing queries”**
 - system alerts analyst of new information as available
 - updates templates if / when appropriate
16. Ability to accept **different types of query input** (e.g. NL, graphical, speech, templates, query by example, models).

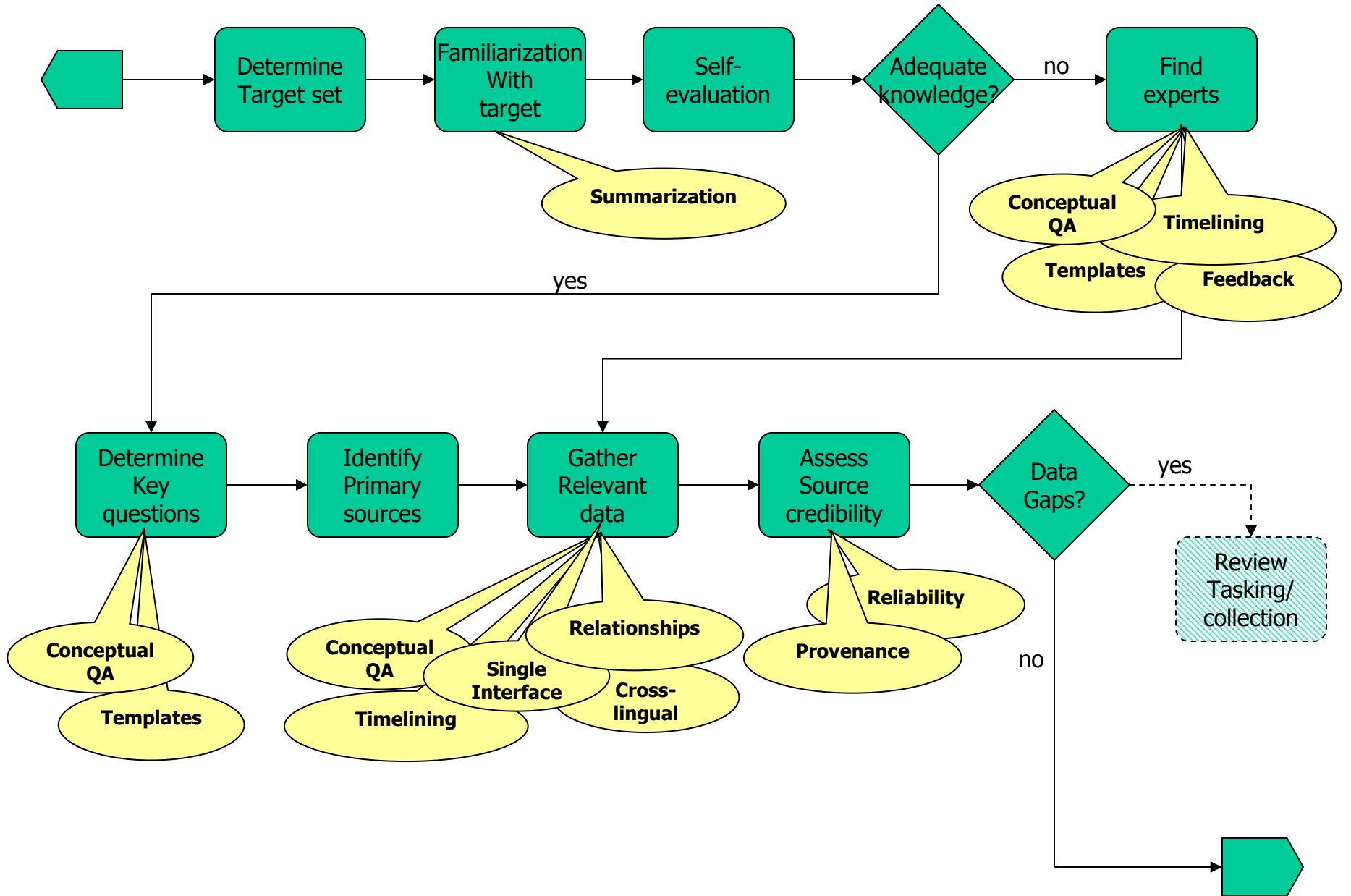
Analysts' Top 20 Requirements (cont'd)

17. Automatic **query re-formulation** when no answer is found
 - System intelligently adapts one variable at a time in the query formulation (e.g. other countries surrounding country in original query)
18. System provides answers from **structured sources** as well as textual ones
 - e.g. bank records, flight manifests, police records
19. System tracks what information has already been seen by the analyst, and provides **only new material** that adds to this knowledge in response to queries
20. A posting area / **collaborative information workspace** for questions & answers of common interest to analyst teams

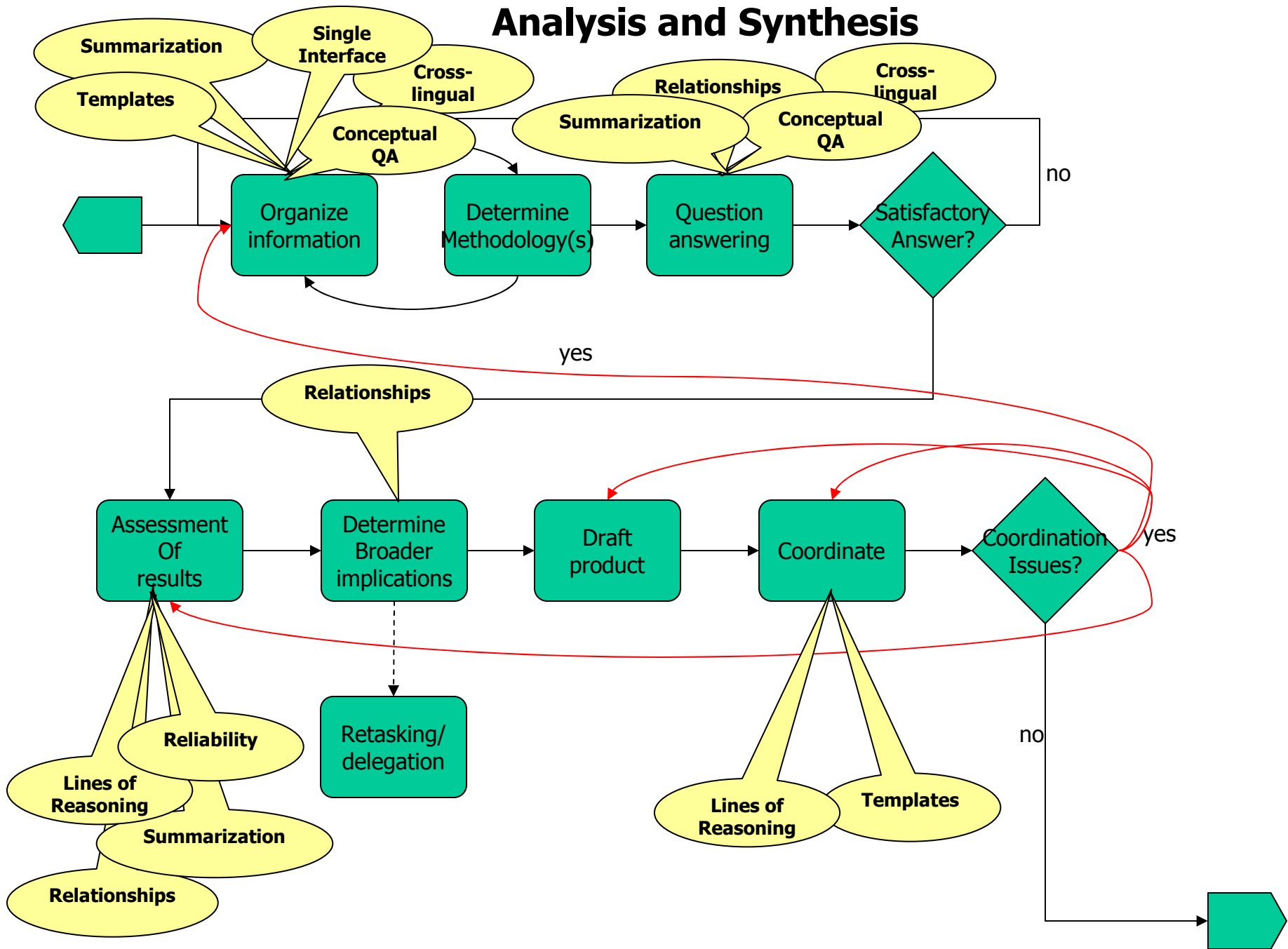
Task Understanding



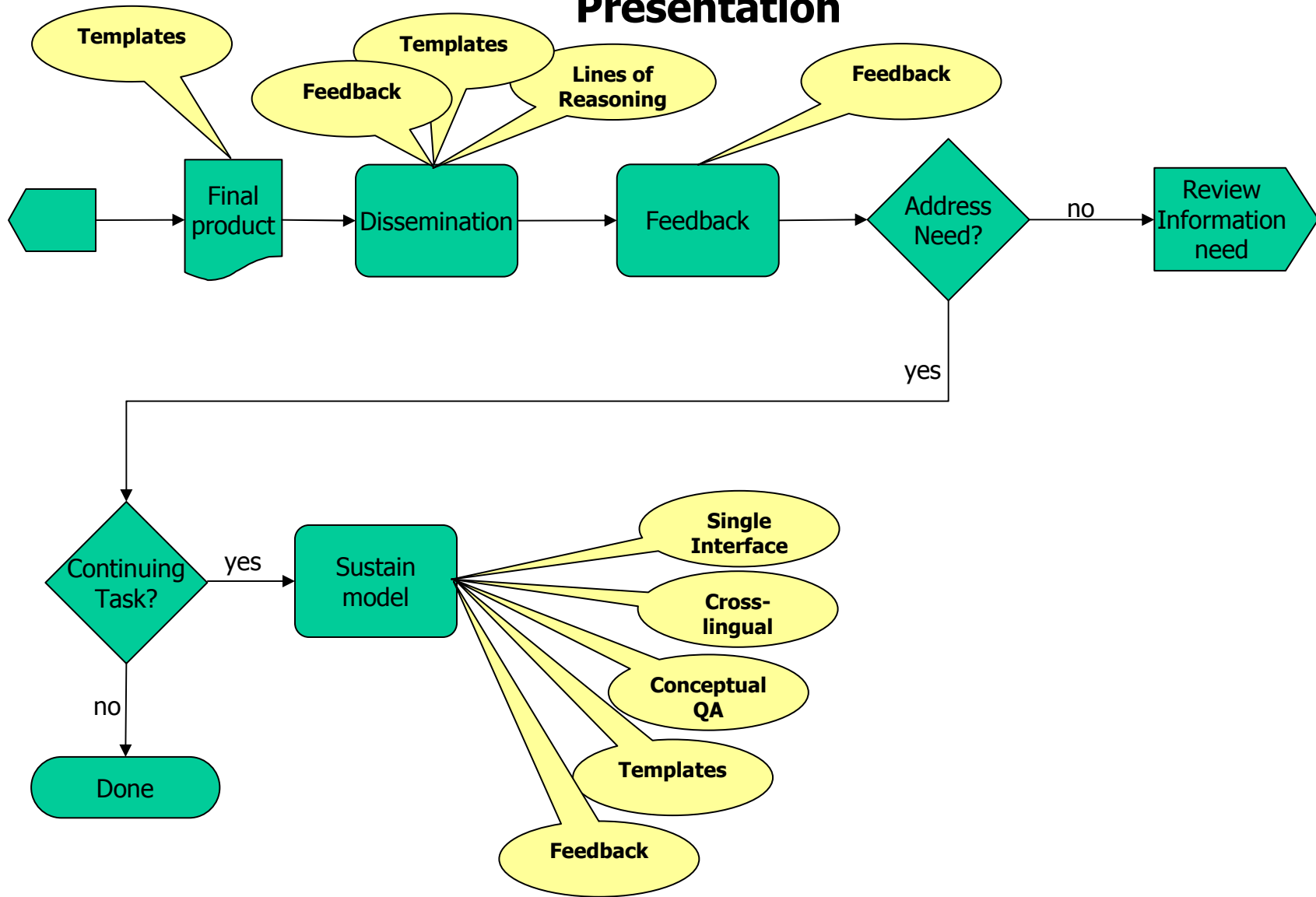
Discovery



Analysis and Synthesis



Presentation

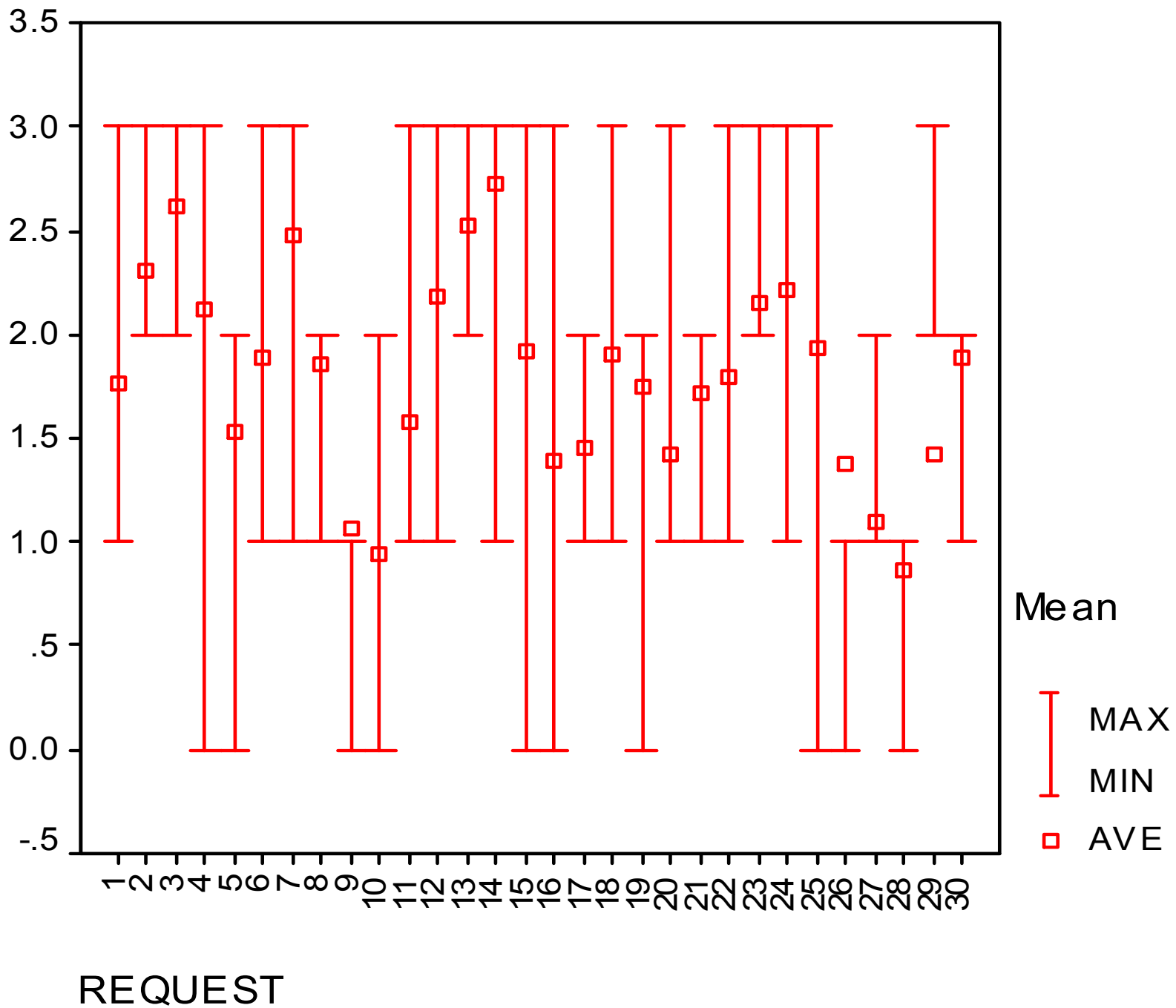




	Phase 1	Phase 2	Phase 3
Goal	To understand Intel Analysts' processes in scenario-based tasks.	To observe the QA capabilities that analysts want to see in systems.	To timeline the QA capabilities needed by analysts to conduct their scenario-based tasks.
Task	Elicit and observe analysts' cognitive & physical actions in accomplishing scenario task via think-aloud protocols and annotations.	Analysts will use QA systems & researchers will interact about how / what technologies might better assist them.	Mapped analyst requirements to QA researchers' estimates of how long it would take to deliver these technology capabilities.
Deliverable	Models of actions and processes that analysts perform.	Analysts' requirements / desiderata for QA technologies.	Research agenda of QA technologies to enable analysts to accomplish scenario-based tasks via their preferred modes.

Time Line Task for Researchers

- For each of the 30 Analysts' requirements, predict how long it would take for researchers / developers to provide this capability
 - A** ---- Already available
 - S** ---- Short term (6 months to 1 year)
 - M** ---- Medium term (2 to 3 years)
 - L** ---- Long term (4 to 5 years)
 - N** ---- Not within researchers' purview to implement
- o Weighted box plots were computed to show range of predicted time as well as average



ROUND 2 – Time Predictions on Desiderata

1. 1 to 3
1.75 Templates of commonly needed information for scenarios which analysts can launch as a query for the system to search and fill in template's slots along with links back to sources.
2. 2 to 3
2.61 System identifies and utilizes **relations** occurring in text, thereby permitting the analyst to query by **relations**, (e.g. *I am looking for **people (X)** who traveled from country Y to country Z.*), not just by topic or entity.
3. 1 to 3
1.58 Answer will indicate the system's line of reasoning / justification for scoring & ranking of responses to question.
4. 1 to 2
1.09 A single system that eliminates the need for multiple interfaces, multiple log-ins, and asking the same question of multiple systems. Answers from multiple databases / systems are merged and ranked.

ROUND 2 – Time Predictions on Desiderata (cont'd)

5. 0 to 2
.94 System indicates where information came from – provenance (defined as place of origin).
6. 1 to 3
2.47 System dialogs with analyst to clarify question, thus enabling analysts to ask natural questions, but ensures that system will clarify question when not clear, as would a reference librarian.
7. 1 to 2
1.86 In response to a question, the system locates individual experts and expert communities utilizing an inter-agency database of who currently is / previously had worked on question.
8. 1 to 3
1.91 Conceptual querying. System represents analyst's need without the analyst having to adjust their query to how the system wants it.

ROUND 2 – Time Predictions on Desiderata (cont'd)

9. 1 to 3
2.18 Ability to understand all temporal references (e.g. *12/8/95, yesterday, current, as of...*); time-stamp the information; timeline it, and therefore; indicate what is the most current information.
10. 1 to 3
2.22 Summarization of results from multiple sources with the ability to drill down into results being summarized.
11. 1 to 2
1.45 Indication of the reliability / believability of a source based on analysts' past experience with the source.
12. 1 to 3
1.89 System deals with analyst's concern for strategic surprises by recognizing unusual patterns of activities and relationships and alerting analyst.

ROUND 2 – Time Predictions on Desiderata (cont'd)

13. 1 to 3 Cross-lingual question-answering capability.
2.73
14. 2 to 3 System recognizes conceptual / logical
2.53 equivalents of responses & indicates why these
responses are deemed equivalent, or if differences between responses, the system will highlight what is omitted from a response, what the responses agree on, & what they don't agree on.
15. 0 to 2 Questions of sustained interest to the analyst
1.53 continue to fire automatically as “standing queries” & system alerts analyst of new information as it becomes available as well as updating templates if appropriate.
16. 2 to 3 Ability to accept different types of query input
2.15 (e.g. NLP, graphical, speech, templates, query by example, models).

ROUND 2 – Time Predictions on Desiderata (cont'd)

17. 1 to 3
1.94 Automatic query re-formulation when no answer is found, i.e. system asks basically the same question, but intelligently adapts one variable at a time in the query formulation (e.g. other countries surrounding country in original query).
18. 0 to 3
1.39 System provides structured answers (e.g. from relational databases), as well as textual responses.
19. 0 to 3
2.12 System tracks what information has already been seen by the analyst, and provides only material that adds to this knowledge in response to queries.
20. 0 to 2
1.38 A ‘posting area’ or ‘collaborative information workspace’ for questions & answers of common interest to teams of analysts.

Reported Benefits of Workshop to Intel Analysts

- *Interaction with experienced analysts from both the intelligence community and academia.*
- *Rare opportunity to meet some of the true experts in the field and gain a much better appreciation of the status of ongoing and proposed research.*
- *“Intelligence Analysis Futures Day” provided a broader view of related analytic activities & issues.*
- *Allowed NSA’s Advanced Analysis Laboratory to advance their organizational primary goal of guiding research to foster and improve analytic tradecraft.*

Workshop Participants here at AQUAINT

Kelcy Allwein

John Burger

David Day

Ulf Hermjakob

Laurie Hiyakumoto

Frank Hughes

Penny Lehtola

Jimmy Lin

Heather Long

Heather McCallum-Bayliss

John Prager

John Prange

Richard Tong

Next Steps

- **Continued content analysis of analysts' processes, preferences, & requirements**
 - How much would be most useful?
- **Move to more abstract models to better reflect **processes** of analysts**
 - More cognitive models vs. physical models
 - Align with various analytic methodologies
 - Analysis of Competing Hypotheses, Possible Futures, etc.
- **Redo researchers' time estimates in more controlled experiment**