

# Building an Online Knowledge Base for Information Infrastructure Protection

Prepared by Michael J. Yacavone for the  
Institute for Information Infrastructure Protection (I3P)  
Dartmouth College, Hanover, NH

FINAL REPORT  
NOVEMBER 2003

This study was supported by grant 60NAB1D0127 from the National Institute of Standards and Technology. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the National Institute of Standards and Technology.



XENIUMGROUP, LLC  
PO Box 828, HANOVER, NH 03755  
P: 603-643-8855, F: 603-619-8929  
W: [HTTP://WWW.XENIUMGROUP.COM](http://www.xeniumgroup.com)  
E: [MYACAVONE@XENIUMGROUP.COM](mailto:MYACAVONE@XENIUMGROUP.COM)

## TABLE OF CONTENTS

Introduction	•	<i>page 4</i>
Key Success Factors	•	<i>page 7</i>
Differentiating Conceptual Platforms	•	<i>page 15</i>
Vision for the Knowledge Base	•	<i>page 20</i>
Content of the Knowledge Base	•	<i>page 23</i>
Functional Scope of the Knowledge Base	•	<i>page 34</i>
Knowledge Base Actors and Goals	•	<i>page 39</i>
Capabilities Roadmap	•	<i>page 43</i>
Appendix A: Synchronous Activities	•	<i>page 45</i>
Appendix B: Research Participants	•	<i>page 47</i>

# Introduction

*METHOD*  
*AUDIENCES*  
*DEFINITION*  
*TIMEFRAME*  
*FOCUS*

## INTRODUCTION

This document describes the requirements for an online Knowledge Base to support the collaboration of academic researchers, industry executives and government policy-makers around the topic of information infrastructure protection.

### *METHOD*

The study was conducted during the Summer of 2003 via telephone interviews with I3P consortium members, research into existing collaboration practices of the community, and observations of online portal and archive best practices. We focused on academic researchers as they are the core constituencies of the I3P, and they currently form a significant majority of consortium members. This focus has the beneficial side effect of targeting a manageable set of requirements for initial software development. The I3P intends to provide the Knowledge Base resources for academic, industry and government participants.

### *AUDIENCES*

This report is intended for three primary audiences:

- Directors and managers of the I3P who need to determine strategy, tactics and staffing to build the Knowledge Base and the Institute;
- Software engineers, information architects and designers who will build the Knowledge Base;
- Consortium members who are interested in the consensus view of how the I3P might help facilitate the collaboration between research, industry and government around information infrastructure protection.

### *DEFINITION*

The term “Knowledge Base” can have various definitions. The I3P Knowledge Base Architect and Manager has defined the Knowledge Base, in part, as an information delivery system with a variety of resources and services available through a variety of tools. Authorized users will have the ability to contribute information to selected areas of the Knowledge Base. This definition is elaborated in more detail beginning on page 17.

### *TIMEFRAME*

We expect that the “Initial” and “Core” visions of the Knowledge Base described in this study will take six to eighteen months to develop and deploy. Some aspects of the Knowledge Base rely on gathering data or meta-data from other organizations and we expect these elements (such as the digital archive) will take

longer to be fully realized. As the Knowledge Base is developed and consortium members provide valuable on-going feedback we expect the content, features and usefulness of the Knowledge Base to grow.

### *FOCUS*

Building a sophisticated Knowledge Base will be a significant undertaking. An overriding goal of this study is to simplify the range of possibilities and bring clarity to the immediate work ahead. While we address long-term possibilities, the focus is on immediate factors that will increase the probability of long-term success.

## Key Success Factors

*DEFINITION*

*COMMITMENT*

*GOVERNANCE*

*INDUSTRY PARTICIPATION*

*SUSTAINABILITY*

*USER ROI*

*MARKETING*

## KEY SUCCESS FACTORS

There are several important factors that will contribute to the success of the I3P Knowledge Base. Following we consider each in turn.

### DEFINITION

Clear definition of the I3P's mission, the goals of the Knowledge Base and the link between the two will be critical to the success of the Knowledge Base. It is vital to communicate a larger, longer-term vision while demonstrating immediate value to consortium members. The longer-term vision presented should be concrete and practical rather than "pie-in-the-sky." The I3P should continuously request input from members to help define the evolution of the I3P Knowledge Base. This report assumes that a feedback loop is in place and is generating useful information.

In early 2003 the I3P had some difficulty defining and communicating its mission to the consortium, largely due to the challenges in bootstrapping the Institute and changes in funding expectations. This study was conducted shortly after a consortium meeting in July 2003 where these challenges were discussed. When interviewed for this study, many people said they were unclear of the purpose or mission of the I3P. When prompted for how a consortium could help, however, they frequently defined an organization very close to the intent of the I3P.

*"That driving force we all need is a place that is a clearinghouse. That is the notion that I still hold firm. We really need a place where both government and research can go and ask 'what is the hard problem in this area?' 'who's working on it and who's not working on it?' 'how do we make some forward motion on that?' When I go somewhere and I ask 'who is addressing this hard problem?' it is not because I want to compete with them or undermine them, it is so that maybe we can team up together. Teamed research is going to move us forward faster than individual research."*

*"I think of facilitation as, broadly, being able to go to the site and look up researchers or workers who are working on this particular topic. This is a form of facilitation as being the first step in me establishing a communication with those researchers. I don't know if this is I3P's role, but you could imagine taking that several steps further. For example, the I3P sponsoring some informal working groups of topics it feels are of critical importance."*

A core function of the I3P is to support synchronous and asynchronous research facilitation. Clearly defining the I3P's mission in the minds of consortium members, and defining the role consortium members play in governing the I3P, is critical to success. The Knowledge Base can help support the Institute's definition with consortium members by providing a valuable facilitation resource.

## COMMITMENT

Paul Thompson was correct when he articulated the challenge of I3P consensus building in his “Concept For an I3P Digital Archive” document:

*“The key challenge for any version of the I3P Digital Archive is consensus-building. For the Digital Archive to succeed, the I3P must obtain the cooperation of individual researchers in academic, government, industry, and the not-for-profit sector. This means that the I3P must build credibility not only with I3P Consortium members, but also with a wide range of institutions outside the I3P Consortium.”*

Some members expressed a lack of commitment to the I3P, largely due to the change in funding resources or “business model” of the I3P. Virtually every person interviewed expressed skepticism that the I3P could obtain enough documents from consortium members to become a credible central archive.

*“Why should we give I3P our intellectual property if I3P doesn’t give us anything in return?”*

This was not a universal opinion, but it should be noted that consortium participation in the research for this study was relatively low (eleven institutes from eight member organizations), indicating a lack of priority in consortium member summer schedules. One dimension of the commitment problem is the following:

*“In computer science there is a hierarchy of places that you can publish and republish. As long as you climb that hierarchy everyone is happy with you. The first place work often appears is in a technical report. It serves the purpose of putting a number and a time stamp on it and making it publicly available in a forum that supposedly will never change. Today many people just stick it up on their website, but that isn’t really citable because links change. A Technical Report is a little more formal in that respect, though of course it is available through the web. Then you can submit it to a conference, and they don’t care if it has been out in a technical report before. Then the next step would usually be a peer review journal, and they don’t usually care if it has appeared in a conference before as long as you have extended it and revised it and made it more complete. So, I can see researchers hesitating if I3P were to publish something – there might be some concern.”*

Based on comments such as the above, we have framed the Knowledge Base as an index of meta-data pointing to resources existing elsewhere, not as a central repository. The Knowledge Base may go further than that eventually, but the I3P must work with consortium members to define the appropriate path.

Several members expressed continued support for the I3P, regardless of the I3P’s ability to fund research.

*“The original idea was that the I3P would actually be doing funding in areas where research wasn’t well funded otherwise. That seems to have changed. I’m not too surprised about that, because it does step on a lot of people’s turf. But, given that change I think [the I3P] should still be there. And the identified research agenda that’s*

*been done now basically just needs to be maintained, updated, continuously pushed. The next step is to try to ensure that research is done to address those agenda items.”*

Some members gained clear long-term value from their participation in the I3P, even without funding opportunities.

*“I don’t mind being part of a group that could help to distribute money even if I didn’t get any of it. I don’t need a graduate student, I’m not a single professor trying to make sure that I’ve got one more graduate student and looking for funding for the summer. I don’t need to get any money out of this in order for it to be valuable. It’s valuable in different ways. For instance, I can see what different people are doing. Would I rather, if the I3P is going to exist – should my organization be part of it or not be part of it? And it’s better for us to be part of it than not be a part of it, that’s how I look at it.”*

Solidifying the commitment of I3P members to the I3P will be critical to the project success. Different members will define success differently. This fact should be accounted for in any communications outreach planning. By framing the Knowledge Base as a meta-data resource, the I3P is demonstrating a commitment to respond to member concerns and react accordingly.

#### GOVERNANCE

Determining Knowledge Base governance policies and making them transparent to the community will be critical. The community will respond to policies and the I3P should be prepared to revise them to accommodate a consensus view. At the same time, it is very difficult for any consortium to govern itself at the micro scale. The I3P must be confident of its ability to lead and coordinate the Institute’s on-going work while creating opportunities for the consortium to provide long-term direction and guidance.

Governance is a well-known challenge in the web portal community. Virtually any published paper or conference report will address the topic. In the corporate world, the question would be phrased, “Who ‘owns’ the portal?” Problems can occur because portals cut across organizational boundaries and impact a large number of people. For example, if the Human Resources department implements a portal for employees to self-manage their benefits package, every employee in the company will be required to learn, use and rely on the portal. But can HR really “implement” a portal? Typically they will require significant support from the Information Technology group and the portal may also interface with one or more outside vendor data systems. In this example, there must be very high-level cooperation and alignment of priorities or else the project will fail – in full visibility of all employees.

Therefore, the governance of portal projects typically follows from overall corporate goals and defines objectives and responsibilities for each department involved in the implementation. Inter-departmental teams manage the project, which can span many months or even years. Frequent communication and sta-

tus reports keep various departments informed. Issues are resolved as they arise or are brought to the attention of management for executive direction.

In addition to general administrative policies, specific Knowledge Base governance will take two forms – content policies and technical policies. At a minimum, the I3P will need to answer the following questions:

*Content policy questions*

- Who owns and is responsible for the information content of each channel?
- What are the criteria by which an information source is included?
- What are the procedures for verifying the integrity of an information source?
- What are the procedures for verifying the continuing quality of an information source?

*Technical policy questions*

- Who owns and is responsible for various technical sub-systems that comprise the Knowledge Base?
- What documents are required for engineers to build sub-systems?
- What milestones or management checkpoints are required to ensure adequate project management?
- What is the process of handling bug reports and feature requests?

The I3P Knowledge Base will probably face on-going governance questions, especially given a consortium model with over 20 complex organizations. For example, because I3P staff will engage in efforts to gather, filter and categorize content – specifically at the request of consortium members – one governance question becomes: Who decides on the filters and categories? It is unlikely that consortium members want to collaboratively decide the details of the filter mechanism, but they will want to know that one is in place, perhaps review the policy, and perhaps have the opportunity to influence the filtering approach. We expect this to occur through the natural course of feedback, and not via a specific working group of consortium members.

Another concrete example arises in the expertise database (see page 23): consortium members want a comprehensive database but not a catch-all Yellow Pages. They want quality listings but they don't want to miss anybody. Who decides what gets in and what stays out? The Knowledge Base can be an effective tool to focus governance questions toward productive resolution.

In some online communities there are voting schemes and other in-place mechanisms for people to register their interest, disinterest or opinion of various sub-systems, content channels or participants. We do not believe these will be effective in this case, based on feedback on the topic of personalization (see page 36). People feel that extensive personalization is too burdensome for the benefit, and we would expect a lack of participation in most content voting or rating schemes.

## INDUSTRY PARTICIPATION

The success of the I3P Knowledge Base will be grounded in the success of the I3P as a consortium. Several aspects of the consortium organization remain in flux, and therefore we expect the Knowledge Base to evolve to meet the consortium needs over time.

One aspect of the consortium organization that is undefined is the role of industry participants. Current consortium members understand the recent history of the I3P, and are supportive of the core I3P mission to bring research, industry and policy together. One typical quote:

*“If you don’t draw that whole triangle [research, industry, policy], you run the risk of not getting the message out to all the people who need it.”*

Several people mentioned the “grounding” effect of working with industry:

*“I think where that connection can be most valuable is that industry has an understanding of what some real world needs and requirements are. Oftentimes, it is an age old concern, academic researchers are too pie-in-the-sky and not in touch with reality. Informed researchers make better researchers, so having industry be invited to give an input into the I3P on a pretty regular basis about what the needs and requirements are would be helpful.”*

Everyone felt that industry participation as associate members of the I3P was important, but everyone agreed that industry participants should not direct the research agenda.

*“I really like the idea of the advisory groups. I would be opposed to industry being full members. That said, industry does have a wealth of information, resources, and so forth, and to cut them out would be absurd. So, some sort of pool of people we can go to, or can help give us advice.”*

Most people defined industry participation in terms of either the CxO-level or senior technical level positions.

*“It is a C level position – CIO, CSO, CFOs – in that range.”*

Frequently, people used the idea of industry submitting “problem statements” that researchers could consider in their work.

*“In addition, it would be nice to be able to find information about these industries or organizations and how their problems can be solved with our technology, or problem statements we can start to address.”*

*“I think that the I3P itself might only want to concern itself with identifying commercial companies that can serve as a transition path for the research, and make that information available to the researchers. That would generally preserve the research focus. Another way to look at it is to allow commercial vendors to provide problem statements and so forth, and basically any researcher that is interested could then have another source when trying to decide how to fund the work.”*

Incorporating industry participation in a way that is satisfactory and beneficial to existing consortium members is essential to cement the I3P mission to be at the intersection of research, industry and policy. This study does not extensively address the needs of industry participants, except to assume their eventual participation in some form. As the Knowledge Base is developed it will provide a concrete focal point for industry feedback and participation.

#### SUSTAINABILITY

Several members commented that a service such as the I3P proposes to offer requires a long-term commitment to maintainability. Regularly updated data, vigilance on data integrity and reliable service operation are all valued and all take on-going resources. The I3P should not promise more than it can deliver. One way to do that is to roll out services one by one to determine the resources required to operate them.

*“In a seven million dollar world, where are you going to target, and what we can put up is certainly a public representation of what we can support.”*

*“I’ve had to make decisions about maintainability. I don’t want to put up a service that people are going to start using and then not be able to support it.”*

The I3P should not over-promise at this juncture. Do not sell the future, but build the present. By constructing the Knowledge Base one feature at a time the I3P is demonstrating the sustainability of the service.

#### USER ROI

One point heard over and over again is that doing simple things well will have significant impact. Grand schemes that cannot be built quickly, that cannot be maintained, and that cannot provide immediate value to researchers have little chance of adoption by members.

Much of the I3P’s work can be facilitated through traditional electronic tools.

*“We produce email very ineffectively in I3P, at least compared to other collaborations of which I am a part, so maybe we could start by doing that. If there were email lists and we were going to use them, if we had something to discuss – I’m not sure we would have anything to talk about – but if we were going to do that they could be archived there, so we can peruse through it so that a bunch of people could peruse back through the archives to see how the decisions were made. If people are taking minutes at meetings those could be posted – that would be helpful.”*

The reason for this short-term focus appears due to an industry-wide problem of historical blind spots.

*“One point that I want to make is that the biggest problem in this field is that we keep rediscovering things. For example, we fixed buffer overflow in the 60’s. I know*

*13P is going to focus on the future and on the present, but I think its biggest contribution would be in also pulling in information from the past.”*

Consortium members also emphasized the need to manage existing knowledge and support resources that already exist.

*“My biggest criticism of what has happened in this field over the past years is that so many people have become aware that this is a problem, and they have been reinventing things that have already been done before, or because they want to be out front and center they are covering the exact same ground that someone else has covered. The institutional resources in the field, and dare I say talent or experience, maybe experience is better, is an incredible waste. That is what really disturbs me the most when I see it happening. Our archive, the one at NIST – [...] doesn’t have a big one – there are a couple like that around and we need to find ways to support those rather than supplant those.”*

Since the desire is for the 13P Knowledge Base to be the go-to place in this domain of knowledge, the Knowledge Base must present a compelling offering that generates browser bookmarks and word-of-mouth email. One way to do that is for the Knowledge Base to focus initial efforts on concrete, high-value, broad-benefit features. We discuss these in detail beginning on page 23.

#### MARKETING

The discussion above points to the importance of strong marketing and communications outreach to consortium members. The 13P can successfully engage in marketing efforts to draw participants into the Knowledge Base and begin building stronger relationships with the consortium community. However, members don’t want to be overloaded with “newsy” email:

*“I would not be interested, for example, in a “SANS in the news,” because I am not interested in whether a SANS is growing, or what impact it is having in the world – what I want to know is what impact it is having on my part of the world.”*

Because Knowledge Base features will roll-out over a period of time, there is ample opportunity to communicate and demonstrate on-going value to 13P consortium members, without annoying them with content-free messages.

## Differentiating Conceptual Platforms

*WHAT IS A WEBSITE?*

*WHAT IS A PORTAL?*

*WHAT IS A DIGITAL ARCHIVE?*

*WHAT SHOULD THE I3P BUILD?*

## DIFFERENTIATING CONCEPTUAL PLATFORMS

It is important to separate the concepts of website, portal, archive and Knowledge Base so that we can discuss the options with a common language.

### WHAT IS A WEBSITE?

“Originally: a computer system that runs a web server (rare). Now: a document or a set of linked documents, usually associated with a particular person, organization, or topic, that is held on such a computer system and can be accessed as part of the World Wide Web.” – Oxford English Dictionary, June 2001.

#### *Problems with the website moniker*

- Sounds simple – everyone has a website; why is this one special?
- Websites don’t typically offer the login or application features envisioned.
- The moniker “website” doesn’t capture the structured nature of information categorization and organization that the I3P envisions.

Short-term, what the I3P is building is a website. Certainly the current web-based presence is simply a website. But this word does not capture the full spirit, mission or aspirations of the project, and so we cannot recommend it.

### WHAT IS A PORTAL?

A portal can be thought of as a doorway to other resources. At the start of this study, some interviewees thought of the portal as a centralized repository of documents. But comments such as the following caused us to re-frame this “central repository” aspect as a separate and distinct archive section of the Knowledge Base:

*“The word portal basically means an opening to other things. Those other things don’t necessarily have to be stored there. The use of the word portal to mean a centralized server that has all the information is a complete misuse of the word and a complete misuse of the concept.”*

This idea of a doorway sounds perfect for the I3P, but unfortunately the portal software industry has expanded this definition to be much larger than simple access to resources. One example:

*“A portal provides internal and external users with a personalized, integrated and secure web-based interface to data, applications and collaboration services.” – Colin White, Enterprise Portals and Web Services Conference, June 2003.*

This industry definition was echoed by many study participants:

*“Q – How do you see a web portal being more than a website? A – I think there has to be some exchange of information, not just we throw a bunch of stuff on the*

*website and you go look at that stuff. There has to be a way for users of the website to inquire further about things, or influence what shows up there.”*

*“The advantage to doing it as a portal instead of just a website is that it means different constituencies could have different experiences there.”*

#### *Typical portal features*

- Integration of data, such as research papers and recent news articles.
- Integration of applications, such as providing a single interface for searching a repository and uploading documents to the repository.
- Integration of business processes and workflow, such as automatically notifying a supervisor that a document uploaded to the repository requires their authorization prior to it being added to the search index.
- Supports a portlet interface to allow others to connect to its datastore, such as exemplified by the Open Archives Initiative data-interchange format.

#### *Problems with the portal moniker*

- The word portal is viewed as a technical term that rings of jargon.
- Portal is a broadly used term that is often defined differently for individual projects.
- It is frequently viewed derisively as a trendy, expensive, content-free, cumbersome and difficult system.

Generally, each person held a different definition of what a portal or an archive would be. Even someone who buys into the portal concept is surprised that they work – since they live on the web they are fragmented by nature.

*“I’ve been really surprised that people have moved away from Usenet into the Web for these things. I think in a way it is a shame because it has fragmented what used to be a single place. You had one portal and you could see a list of all [the newsgroups], you can’t do that with web portals or web communities. Because there was a standard protocol I could choose my own newsreader client. They still exist, but it surprises me that so much has gone away from that.”*

This quote summarizes the problems of integrated portal applications:

*“Even here at [...] we have these sort of portal packages, in the industrial strength that are available for use in the university. They tend to die out for lack of content. Because it is more effort putting content in, the researchers just don’t have the time to do it that way and the search tools aren’t good enough. I will often use Google to search for things here at [...], rather than using some portal’s own search tools. It works just as well, it’s a familiar interface, and why bother otherwise. And it finds more things, even though I know what I am looking for is here at [...]. I think you are going to run into the same sort of problems. If you say we are going to make this nice, expensive portal it is going to die for lack of content because the people that you want to be providing content are going to find that all of a sudden you have made their lives harder. Making it easier for the content users at the expense of the content*

*providers in the research community means you don't get content. In the research community people know how to find the content well enough."*

But, even people who expressed skepticism about building a portal still had uses for online tools:

*"I guess if there is a member's only part of it and that could have all the relevant documentation that is produced, address list, information about where the next meeting is, and things like that, I think that would be enough for me right now."*

As defined by the portal software applications available off the shelf, a portal has little to do with the goals and capacities of the I3P Knowledge Base project in short- and immediate-term development, so we cannot recommend calling the project a portal.

#### WHAT IS A DIGITAL ARCHIVE?

A portion of what the I3P wants to build is indeed an archive. Archives have a fairly specific definition in the field.

*"The National Archives definition of "ready access to essential evidence" fits nicely into the I3P mission. How that happens is a process of collection, control, preservation, and retention of records about objects. In working with digital archives, that definition has had to expand, but the emphasis is still on management, control, preservation, and retention of records about digital objects."*

Archives also come with the burden and opportunity of national standards and interoperability.

*"One thing to keep in mind are the credibility issues related to building the archives, i.e. adherence to standards and digital archival protocols, will have significant implications. If we want to go to NSF for funding to grow the archives they will want us to conform to national standards. Additionally, a number of I3P members are already building archives that conform to those standards and we may want to join forces with one or several of them."*

Much of what consortium members want from the I3P, especially in the short term, is unrelated to an archive and the term would be too limiting. Even when phrased as "portal and archive" the combination still does not capture the goals that are elaborated in this research. While we do believe that the digital archive has a place within the I3P Knowledge Base, it is best developed later in the program after other offerings are deployed.

#### WHAT SHOULD THE I3P BUILD?

Based on the above input we feel that the I3P should re-conceptualize the asynchronous online development effort as a Knowledge Base. The term Knowledge Base can have a number of different definitions. The I3P Knowledge Base Archi-

tect and Manager has defined the Knowledge Base, in part, as an information delivery system with the following characteristics:

- The system will deliver a body of information to users under the broad category of information infrastructure protection;
- The system will allow the user to find information contained within the Knowledge Base through a variety of approaches, eg., by topic, author, institution, type of information, etc.;
- The Knowledge Base will contain resources, services and tools in a variety of formats;
- Authorized users will have the ability to contribute information to selected areas of the Knowledge Base.

Rather than thinking of the project as a collection of web pages, or a repository of documents (or other potential content), or a doorway to other resources, the I3P Knowledge Base should be thought of as all those things, packaged in such a way as to deliver an integrated, topic-specific resource for the information infrastructure protection community.

*“What we would like to see as members of the consortium and as users of whatever portal they build is information-sharing that helps us direct the research or branch out and talk to other people, be they researchers or government institutions, that might be able to utilize the work, or even industry people that might be able to use the work.”*

A Knowledge Base captures the idea that what the I3P is building is more than web pages, is more useful than a portal and is collecting more than documents. It is really a web-based application that will help content editors gather, summarize and link useful and valuable resources to inform constituents and facilitate their work, and their collaborative work together.

# 4

## Vision for the Knowledge Base

*INITIAL VISION*  
*CORE VISION*  
*EXPANDED VISION*

## VISION FOR THE KNOWLEDGE BASE

The overall vision for the Knowledge Base is derived from the research to define the specific content, scope and actors of the Knowledge Base system.

The world of possibilities for applying technology to collaborative communities of practice is infinite, but resources are finite. While not intending to limit the long-term breadth of the I3P Knowledge Base, immediate one-foot-in-front-of-the-other progress is required. Exceeding the expectations demonstrated in the following quote would go a long way toward building credibility and commitment for the I3P in the minds of consortium members.

*“I would think getting the portal decided and what constituencies are going to be in the first phase – I would expect nothing less than a year. I think that may in fact be optimistic, because of the number of people involved in this process. I think they need to move much quicker than that, but I think it is probably going to be a year before they show it off.”*

To effectively sequence the work ahead, we have defined Initial, Core and Expanded vision statements for the Knowledge Base. Consortium members will provide feedback on the Initial and Core Visions as they are realized, and we would expect the Expanded Vision presented here to evolve significantly in response to such feedback.

### INITIAL VISION

In the Initial Vision, consortium members want an expertise directory, including contact information on colleagues and their projects; pointers to funding opportunities; and information on related topical news and events.

*“Being able to act as a conduit where people learn who’s doing related things, and connecting, making those connections. Just pointing out who’s doing what and how it is related.”*

No frills, straight information, delivered conveniently and without extra work on the part of the researchers. While researchers eventually want the I3P to deliver more than just this Initial Vision, it is important that the I3P Knowledge Base deliver some value sooner rather than later, and everyone could agree on the value of this Initial Vision.

### CORE VISION

The Core Vision brings in the idea of creating working groups and providing online space for their work.

*“One vision of the portal would be less emphasis on the work output and more on facilitating the work of the group offline. And that’s the model we see elsewhere. And when we run these things it’s not always pretty, but they have the mail archives, the*

*critical documents for the group, they have a schedule of some kind. Whenever there is data to be shared that's where it goes. They are typically password protected and it's kind of one stop shopping for the group. There's links to other things, even if we are not going to store it we will link to it if it is important to the group."*

The Core Vision presents the I3P Knowledge Base as a vibrant center of resources and research activity on information infrastructure issues. It assumes a fully-functioning Institute where there are productive meetings in person, there are a number of working groups moving projects forward between consortium meetings and a central online hub is being built that connects academic research, industry and government.

#### *EXPANDED VISION*

An Expanded Vision for the Knowledge Base is that the I3P is creating original work, and publishing that work to a wide audience.

*"I think what the I3P could really add as the value here is the meta-knowledge, not just the facts, but how to use them in context, how do they fit into these things. This material gets published as best practices, policy and procedure. This is where I would think they would have a number of interesting research areas, not just in CS, but also in policy, in business practices."*

This view was heard from I3P, ISTS and the Dartmouth community, but never from other consortium members. The Expanded Vision for the Knowledge Base is still possible, but it will require success in delivering on the Initial and Core Vision components first. It could take several years building the relationships required to deliver the full scope of the Expanded Vision.

The I3P should focus on building the Initial Vision while considering the Core Vision, and put the Expanded Vision on hold during the immediate deployment planning.

## Content of the Knowledge Base

*TYPES OF DATA*  
*EXPERTISE DIRECTORY*  
*FUNDING DIRECTORY*  
*EVENTS CALENDAR*  
*I3P MEMBER PROFILES*  
*WORKING GROUPS*  
*DIGITAL ARCHIVE*  
*LITERATURE SUMMARIES*  
*CONFERENCE SUMMARIES*  
*RESEARCH DATA SETS*  
*ONLINE JOURNAL*  
*PRACTITIONER INFORMATION*  
*SPECIAL PROJECTS*

## CONTENT OF THE KNOWLEDGE BASE

By way of introduction, we should consider what types of data we can expect in the Knowledge Base. Generally, there are three broad types of data.

### *Unstructured data*

- Documents
- Media assets

### *Semi-structured data*

- Tagged information such as meta-information about documents, or documents with meta-tags.
- XML data

### *Structured data*

- Database data

It is important when considering each content channel to know if there will exist unstructured, semi-structured or structured data access. The greater the structure the greater the functionality that can be provided.

## EXPERTISE DIRECTORY

There is extremely strong interest in an expertise directory, and it provides an excellent cost/benefit ratio for research, industry and government participants. Described over and over again, consortium members wanted to find out who was working on what, where, and what agencies funded the research. This would obviously be an equally valuable directory for government and industry to find appropriate people and institutions for research and collaboration.

*“One of the things that is really bugging me, is a way to learn what other people are doing and the kinds of problems they have, and the attacks they are facing.”*

*“What I always wanted, what my researchers always wanted, and even what many of my government customers always wanted, was to know “who else is doing research in this?” You sit down and write a proposal to the government and they want to know related research areas, you really want to be able to understand who else is doing research. It is a very important question and it is something that frustrates me tremendously. I think we are wasting a tremendous amount of dollars in duplicated effort and no one knows enough about who is doing what in what area.”*

*“Maintain an index of groups working on particular topics that have been identified in the research directions document. Speaking for myself, that is something that would be quite useful, not only to be able to see the topics, but to know who in their own words is working on each one.”*

*“Pointers to who is doing research in what areas, not just inside the consortium, but outside the consortium. Probably break that down initially by the research agenda. Initially there would have to be an effort to try to build this, maybe search to find some things, and eventually it would become self-perpetuating if it is out there. It would be especially beneficial to move beyond separate disciplines. It would not be just an expert’s guide. It needs to be more than that, experts, interests.”*

*“There would be indexing by people, and also by groups. This is the group at the [...] doing XYZ. Both of those would be useful and interesting ways to approach it. Sometimes you know the name of the group, but you don’t know where they are, sometimes you know the name of the school, sometimes you know the name of the person, sometimes you just don’t know. It is hard to keep those things up to date, but it would be a useful service.”*

Another feature of the expertise directory is contact listings for government policy makers. Developing organizational charts of where people are and what they are responsible for has very high value to the researchers. People change job responsibilities frequently and they are often difficult to track. If the I3P regularly updated this funding contact directory it will provide another reason to visit the Knowledge Base.

*“What you may find is that some of the funding opportunities are targeted and not public, others are very public, like the TC or digigov stuff from the NSF. Those would be good. Also a list of contacts at the various agencies. For example, I’m doing some stuff with the local county recorder, and we suddenly realized that the Department of Homeland Security might have an interest in some of the stuff we are doing. I didn’t get a contact until yesterday when I was at a meeting of the National Science Foundation in Washington.”*

There was concern that the I3P would need to maintain the integrity of the expertise directory. An important challenge is to minimize bias in the selection and listing of people, institutions and categories. There is a desire for high-value information, and everyone understood that this means filtered, edited or otherwise reviewed information.

*“It would also need to be vetted to some extent otherwise it becomes a sales tool rather than a resource. You can’t just let people post their own or you get a few companies that have products to sell posting everywhere.”*

*“If you start listing individual companies or individuals it could get a little dicey. If you kept it open to have anyone listed there would never be a question of bias. Then it becomes a yellow pages.”*

Developing a governance policy to guide inclusion and exclusion will be critical. This policy can be revisited and revised as required, but having some published guidelines will help the consortium understand what the expertise directory is, and will also, by nature, solicit comments on how to broaden or narrow the directory contents.

There is also a sense that the directory could be helpful for the media to find people willing to speak on timely security issues.

*“It used to be I wouldn’t even talk to reporters, but I’ve realized there is so much misinformation out there that it is actually more helpful to say a little bit and say it right. Even if they butcher it, at least you are getting the information to the reporter and they begin to think about it. When the opportunity arises it is important that the reporter get accurate information.”*

There is a small twist in providing broad access to contact information. Most people want their personal data protected to minimize nefarious uses, however, the media are well-known for working on tight deadlines with demanding turn-around schedules (i.e. less than two hours). It is possible that an outreach effort to pre-register the media will help identify the I3P and the Knowledge Base as a valuable media resource.

The I3P could also provide a website “Media” link that points to an I3P contact person. The I3P representative would work with the media contact on an immediate-turnaround basis to find the appropriate research contacts and provide several possibilities. The media representative could then contact the identified researchers directly.

#### *FUNDING DIRECTORY*

There is strong support for the I3P to track BAAs and other funding opportunities. Everyone acknowledges that this is a difficult and time-consuming job, but that it would have very high value to researchers. Centralizing the effort to collect this information will maximize the cost/benefit ratio among consortium members.

Categorizing the information in terms of funding agency and funded projects has value to the community.

*“What might be of more interest are [listings of] specific funded projects. For example, we will be doing a project on analysis of vulnerabilities for the National Science Foundation. That is my main research area, but [seeing] that I have a project [at NSF] will really help [others] identify what I am doing.”*

People were emphatic that the funding directory contain links to the funding documents and that they not be stored in an I3P database. The reason is that the BAAs change frequently, and any stored document would shortly be out of date and they will have to hunt for the BAA anyway. Much better for the I3P to simply link to it and then they have the best of both worlds: Aggregated funding information in one place and direct access to the critical documents.

## TOPICAL NEWS

Frequently changing content keeps people visiting websites. Topical news is the most frequently changing content available. Three types of news feeds have been identified:

- Security news
- Funding news
- Policy news

In addition to website postings, email delivery is preferable for many people.

*“I prefer any content to be pushed at me through email, that’s my style, because then its all in one place.”*

It is possible to include several different email news channels. In addition to Security News, Funding News and Policy News, there might be email newsletters for Event News (new material added to the calendar – this in itself could be several different content feeds) and I3P Institute News. There is a lot of competition for readership and attention by news sources, but if the I3P can be responsive to the narrowly targeted needs of its members topical news will be highly valued.

In the near future, subscribing via RSS (Rich Site Summary AKA Really Simple Syndication – an XML format) is likely to be popular as well. This is not currently substantiated by consortium member comments, but is intuitively obvious watching the explosive growth of “weblog” technology, largely fueled by RSS subscriptions. Implementing RSS feeds is trivial, technically.

## EVENTS CALENDAR

Members feel that an integrated domain-specific event calendar is a useful resource.

*“That would be a valuable service. Right now no one is really doing it. If you are out in the field all these things show up over email, because you are on all the email lists, but otherwise you have to go to IEEE see what they have coming up, go to USENIX, go to ACM and so forth.”*

In addition to the event listings, members also mentioned that the calendar should include the deadlines for paper submissions, committee decision deadlines, final paper submission deadlines, registration deadlines, etc.

*“For example, do not say ‘here are the ACM events, here are the IEEE events.’ Put the calendar up and then put links to the call for papers and so forth. And also, if you are going to do that a good idea would be to include dates for conferences and due dates of special papers, and also have due dates and registration dates, so people know when they have to have stuff in.”*

People do not want the information pre-sorted in advance, but instead want to filter in various ways during their real-time search. There should be a single

interface, designed with filter criteria such as academic vs. commercial vs. trade-show conference, as well as filter criteria on topics such as security, information, database, trust, eight agenda areas, etc.

*“The usual ones are academic, professional, trade show, that sort of thing. Also be aware that different universities have different standards. For example, for some people Oakland would be considered academic for all, the applications conference would be considered academic, for others it would not be academic. What you might do is draw a distinction and say which ones accept submissions for papers, which ones accept position papers only, which ones don’t accept anything it is just presentation, colloquial, things like that.”*

Some members are concerned about keeping calendar content valuable:

*“I’m nervous about that because the field has gotten so hot over time, and on any given day I can open my mailbox and have dozens of announcements and only 10 percent of them have any merit. There are a lot of so-called conferences and workshops that are put on by commercial or conference houses that just do nothing but make money by putting on conferences. If you tried to make a calendar, how would you screen out what I would call the garbage, versus the real stuff?”*

Clearly, an editorial governance policy will have to be developed by the I3P staff and reviewed by I3P consortium members to find the right balance between inclusiveness and focus. By supporting multiple filter criteria users can choose what is appropriate for their requirements.

### *I3P MEMBER PROFILES*

No one reacted poorly to this possibility, but it was considered obvious. Typical comment:

*“Sure, gotta have that. Not too much value but it would be missing if it were not present.”*

The profiles are distinguished from the expertise directory in that the member profiles contain just I3P members, and the expertise directory contains entries for anyone working in the field that falls within the inclusion policy.

### *WORKING GROUPS*

As mentioned above, when work is funded by the I3P, some people desired a central place to post documents and information related to the work. Also useful is the ability to use an email listserv for project communication. Most people have that sort of facility available at their organizations, but setting it up and managing it is sometimes problematic. Embodied in this view is the idea that the I3P might provide working group assistance such as setup and management of listservs and webspaces. Although listservs can be time-consuming to keep active, in the context of a specific project working group the listserv will be used to the

extent that there is shared work among participants. In other words, it is a potentially helpful tool, not a required aspect of a working group.

### DIGITAL ARCHIVE

There are two pieces of the Digital Archive – the database of meta-records (which can be thought of as bibliographic records) and the repository of objects (documents, images, etc.). The repository will have a collection development policy (what is retained and why), whereas the meta-record database will link to objects both inside the repository and objects held elsewhere. The repository will contain only objects created and “owned” by the I3P. An archive repository implies a level of preservation and maintenance that the I3P will consider only for information it owns. The primary reason for an archive is preservation of the object. This remains consistent, be it for physical objects or a digital image. Only the methods of preservation differ.

Part of the complexity of the Digital Archive has to do with an existing and finely developed national and international structure for archives. The challenges the I3P will face will have to do with the policies and cooperation required for harvesting the meta-records, not the technical standards themselves – those have been in place for quite awhile – and coming up with a preservation plan for electronic information.

The standard for creating meta-records is the Dublin Core – a national standard aimed at smooth exchange of meta-records created for objects held in repositories and archives.

The appropriate scope of the archive is highly dependent on the collaboration practice of the community. Initially the I3P evaluated creating a repository-style archive. Comments such as the following have influenced the change toward a meta-data archive of external resources:

*“Q – What would encourage you personally to add to this archive?  
A – Money. Most people, a lot of people, that will be the answer, but if there was a sense of community in the consortia where the members really benefited from each other’s presence and there was a certain quid pro quo where people feel they need to give back based on the useful help they already received, I think that would be a big motivator.”*

*“The way these different archives have been received has differed quite a bit, so the idea was it depends on paper base or previous culture that existed in the community as to how the e-archive will be received. For example with physicists they had tended to communicate pretty openly prior to publication, where as some of the biomedical sciences there was sort of an opposite way.”*

In the end, people were clear that a central resource that contained meta-data and links was more valuable than a repository that had only a small collection of documents.

*“A good model to pattern after would be CERIAS, although CERIAS made the mistake of bringing everything in instead of providing links.”*

The copyright and pre-publication issues are foremost in the minds of consortium members:

*“I think you would have trouble putting copies of the papers there. On the other hand, I don’t think you would have trouble putting abstracts and links, but by the time you add up copyright issues, and ownership issues a lot of people would not want a copy of the paper to be publicly accessible on a third party service.”*

Some people thought a possible short-term opportunity would be to include government reports, which are not covered by copyright:

*“Another thing that might be good is government reports and stuff like that that are not copyrighted, putting links to them and somehow classifying and organizing them. For example if I have a question about the DMCA and what papers have been written that discuss problems researchers have I could look them up.”*

The archive needs a more detailed definition and data model, and that work is underway by the I3P Knowledge Base Architect and Manager.

#### LITERATURE SUMMARIES

Literature summaries seem interesting to several consortium members. These are defined broadly, from table-of-contents summaries to original survey articles, but there is an aspect of “keeping track of the field” that seems within the I3P mission.

*“Ingenta abstracts a million different journals and conferences, and once you have found a journal that you want on their website you click, ‘I want to subscribe to that one’s table of contents.’ Every couple of days I get an email message from them with a new table of contents. I have about 30 or 40 journals that I like to track.”*

*“Something I think would be really valuable is a service that was focused on I3P-type things that regularly scanned the technical literature, and allowed you to subscribe to a certain subset of that. Especially if it came with a solid link to the paper on the publisher’s or author’s website, so you didn’t have to spend another 30 minutes digging up the paper. I think that is something that the I3P research community would find extremely useful.”*

*“Beyond knowing topics of importance, it usually boils down to which conferences are publishing computer security material. What computer security papers have been published in those conferences; what are recent presentations. Of course, maintaining such an index is a relatively time consuming, but if it can be done it adds a lot of value for the community. It dramatically improves the accuracy and speeds up researchers review of prior work, which in turn dramatically improves the projects they are putting forward to the government and other entities.”*

*“Lots of technical journals, like the IEEE, do survey papers. Q – Would the I3P have any mission in creating those surveys? A – I think so. I think analysis of open source information in that field would be very valuable. The IEEE does have security and privacy magazines, but it doesn’t necessarily deal with the level where you would be talking about infrastructure protection. That requires insider information that is not necessarily academic level, so someone needs to step in and review that open source information and analysis. That role is open, so if the I3P puts that on their list I think it would rise to the top of what people would want.”*

Our recommendation is that the I3P should investigate the resources required for literature summaries. It is possible that a relationship could be negotiated with the appropriate publishers, or that I3P staff could abstract information from a core set of published material without excessive effort.

#### CONFERENCE SUMMARIES

Conference summaries get a mixed report. Some people think they are valuable for capturing the un-official aspects of conferences. Others feel that the quality is too variable to make them valuable.

*“Another thing I found really useful, is a professional organization, called USENIX. One of the things they do is put out a conference report. They get graduate students to put out summaries on the different conference sessions. This is different than the abstracts you would read in the proceedings. The nice thing about this is they talk about the presentation, the questions, the answers, the buzz around the presentation that wouldn’t be written anywhere else. You definitely get a lot of variation in the style and quality of the write-ups, though I still find these a good way to get an overview of the conferences.”*

*“We have people who frequently did those [conference summaries] and sent them around. They were interesting, but they are primarily of interest to a fairly focused group. They are going to come from someone’s viewpoint. Unless your viewpoint is the same as the author’s it tends not to have that much value. The value of sending a graduate student to a conference to get that write up on a web page – I think the funds could be spent better a lot of other ways.”*

We cannot recommend conference summaries as a core part of the Knowledge Base at this time. We do not believe they should be removed from consideration, but they form a possibility for the Expanded Vision of the Knowledge Base.

#### RESEARCH DATA SETS

People viewed this as valuable, but a long-term effort – no one thought it could happen for several years. Other than the Lincoln Labs data, which is synthetic and not from an actual network intrusion, no one knows of any other data to archive. The problem is that this data usually contains, at the very least, proprietary information. Sanitizing this data is extremely time-consuming, must be

performed by the data supplier due to the nature of the contents, requires the direct participation of a senior researcher to insure that the data is not corrupted during sanitization, and often removes much of the value of the data set.

*“I can almost guarantee you no one would be willing to give you real data. What they might be able to do is give you a pointer to sanitized data, but there are a lot of legal issues, and privacy issues. In particular in the state of California, it took three years to get permission to monitor our own networks, because of the privacy rules.”*

The research data sets should, at the earliest, form part of the expanded vision for the Knowledge Base.

### ONLINE JOURNAL

This was viewed as impossible for the I3P to create in the short-term. Everyone interviewed considered an online journal much harder to create and operate than an archive. An online journal takes years of dedicated effort to create a reputation and a flow of incoming submissions. While the idea of publishing, for example, an annual journal of the I3P-funded work was well-received, this was not considered an online journal, but something along the lines of an annual report.

### PRACTITIONER INFORMATION

Several people, many of them at Dartmouth, suggested FAQ's, How-To's, product selection guides, book lists, etc. as valuable content types.

*“The scenario I can imagine occurring is someone who wants to find out, ‘how do I make my place more secure from cyber attacks? I need to find some experts, some best practices, or I need to find some books, or some courses, or I need to send my staff to this seminar.’ The practitioner needs that kind of information. If they come and all they find is a bunch of research papers they are not going to find it very useful. I can certainly see some content targeted at practitioners.”*

Another perspective was to offer something of a Q&A service, to aggregate the most important questions.

*“You could have people submit questions for weekly response by the I3P. Have some technical team that does some research in the types of things that people are interested in.”*

Others reacted negatively to the extensive resources required to provide quality practitioner information:

*“I think you would need enormous resources to answer questions like ‘how should I configure my firewall?’ unless that was done with FAQs, or on a weekly basis, and you dedicate a technical lead to answer the top ten questions. From a public side point of view, I do think you have to limit how and what kinds of questions you are going to answer.”*

People with experience in this area were skeptical that the I3P could be fair, balanced and keep up with the changing marketplace.

*“Probably about six years ago, as a class project a number of my students ranked some of the tools in terms of ease of use or value for an administrator or various criteria like that. Then we averaged all those things together and tried putting ratings on some of the tools. What we got back as feedback was some people were outraged at the rankings we gave, the fact that we weren’t able to do all the tools in the category, some tools didn’t get rated, and so people thought that meant they were poor, things broke between releases, so the old ratings weren’t proper. As a result it became stale rather quickly. We came to the conclusion that trying to do any kind of ranking like that would require a full time staff of three or four people, just to keep up.”*

Best practices and policies got mentioned both as practical information and as education and guidance tools for wider audiences.

*“The other thing is you might look at having different policies or collections of policies. EFF has some, there are various other places, but links to policies that universities have set up, linked to good books, or popular media, and that sort of thing, to make it amenable to not just researchers, but newspaper reporters and so forth. Hopefully that would serve to dispel some of the misinformation floating around.”*

Practitioner-level information may be part of the Expanded Vision of the Knowledge Base. Other than collecting pointers to best-practice policies, useful books and other popular media sources, it will require significant resources and should not be undertaken without long-term commitment.

### *SPECIAL PROJECTS*

Some ideas were not easy to categorize, but could be important work that is within the I3P purview and resources. The following could be considered a “special project” request:

*“One thing that would be really cool is to get an oral history from many of the old timers. I’m sure they would be more than willing to participate in writing a history of computer security, but something like that would really show how some of the older mechanisms we had can be moved into modern equipment, or how modern equipment can be rearranged to support them.”*

The I3P should gauge member interest in special projects on a case-by-case basis. In some situations there may be particular topics, skills or time-sensitivity to which a consortium member is well-suited. If the project falls within their mission, the I3P may choose to seize the opportunity in these cases.

# Functional Scope of the Knowledge Base

*LINK TO VALUABLE RESOURCES*

*SEARCH*

*BROWSE*

*SECURITY OF THE KNOWLEDGE BASE*

*PERSONALIZATION*

*TEAM COLLABORATION*

## FUNCTIONAL SCOPE OF THE KNOWLEDGE BASE

The functional scope of the Knowledge Base refers to the complete set of features available to all possible actors. The required scope is dependent on the collaboration practices of the community. Our hypothesis of the computer security community is that they are willing to collaborate; the economic gain is not a pressing issue; there are no formal structures which exist to collaborate; there exists a culture of independence that lends itself to one-on-one communication; this is a social group; many people spend too much time online; many people travel too much already.

### LINK TO VALUABLE RESOURCES

An overriding concern is that the I3P should link to existing resources and not re-invent the wheel. Some of this certainly stems from the historical blind-spots discussed earlier.

*“Maintaining your collection, and there are probably some things that are worth maintaining a collection of, but for the most part you do that and you end up with a copy that goes out of date. You are better off just giving a pointer to the source right off the bat. With a one liner or maybe a paragraph summary, so they know if they want to track to that source. That is more beneficial to the community, and it is lower cost.”*

One comment on linking, frequently heard, is that distributing the maintenance work is an advantage. When the work is centralized in a single system it often becomes too cumbersome to use.

*“The reason they tend to be at someone’s personal site, as is done with these [ieee-security.org], is because it is easier to do it that way than to use existing commercial grade resources that do exist and are available for these purposes. If I had to do something at Dartmouth I would say forget it, put a link to my site, because that is a lot easier. I think you will find that where someone has done something elsewhere, look to put a link to it. That is the whole power of the web.”*

In summary, at the present time consortium members do not see a path to the I3P becoming “the” central archive, so collection efforts should occur over time, gradually building meta-records for the archive. Further details are discussed on page 28.

### SEARCH

Perhaps the most important quote in the entire report is this one:

*“I don’t think it needs to be extremely complex. The main thing is to be able to search just like you would be able to a card catalog. In my head it is not something really fancy.”*

That said, searching text content is a well-known difficult problem. It is also a central part of what people expect to be able to do at a Knowledge Base, portal, archive or website.

Generally speaking, the Knowledge Base might include content search, parameterized content search (“advanced search”), citation search, and integrated search of the Knowledge Base and the web. In addition, there is meta-data searching of external archives.

One promising approach would be to use the Google search appliance to create a topic-specific search resource. The product is a hardware solution that can be pointed at internal and, with permission, external HTTP resources. It then provides Google-quality searching of the targeted resources. While this approach requires further, more detailed evaluation, it could provide a very high-quality initial search solution. The Google brand would assist in validating the perception of the search result quality.

## BROWSE

Browsing content by categories is a very useful tool that requires taxonomy development. In this case, the I3P content will include everything described earlier, so filtering mechanisms will be required to reduce the browsing to the desired content types.

*“For me it is taxonomy, knowing who is doing the work, where the work is, who the funding agencies are, hooking up the right people between the different funding agencies. Very challenging to develop, but I do think a taxonomy would be excellent. In my experience, even in smaller organizations we all had our own interpretations of terminology.”*

*“I would also like to do finer granularity in the categories, but that requires someone with some expertise to go through and develop the classification scheme. We actually have a couple of linguists that work in the security ontology field here, so that is not a problem, but we have to then be able to use it in an appropriate way, declassify the things.”*

Each content type will have different browse entry points. For example, the expertise guide might include:

- Browse content by institution
- Browse content by author
- Browse by funding agency
- Browse by project name (grant title)

## SECURITY OF THE KNOWLEDGE BASE

Most people did not perceive a critical security need for the Knowledge Base. Everyone agreed that there should be basic password protection, and we can

envision scenarios where a finer granularity is required. However, the problem with trying to build a secure system can be summed up in the following quote:

*“To put it bluntly, I don’t know your site, and unless you can convince me you have top-notch people running it, I’m not going to trust it. What I would suggest is you have pointers to my website and if I don’t want someone in there I’ll protect it. That way, again, you are not taking on my control, each group can exercise their own controls. Also, if I am dealing with proprietary data, if it is on a local system, normally, I will have that system locked down tight. If there is an accidental breach, the company I am doing the work with will be aware of exactly what I’ve done. If there is a breach it will be clear there was no violation of duty, or we didn’t know about this, why didn’t you tell us about this, we share the blame. On the other hand, if I give it to you, the company doesn’t have that same assurance, because they won’t be involved at all in protecting your site. I would be very reluctant to do that.”*

Basic user-level security will be required. In some cases there may be a need for protected-item security.

#### PERSONALIZATION

The traditional view of portal interfaces is that the user can choose modules and the organization of those modules for their own personal view into the portal content. This idea of a unified user interface to multiple services has strong pull in corporate environments where there are multiple operational systems. However, given the tightly targeted nature of the I3P Knowledge Base that evolved in the Initial and Core Visions, people responded poorly to the idea of a sophisticated interface, especially one they had to learn.

*“Just give me links. As a person, as a human being, I have too much of that [customization]. I have to go to my payroll website, my bank website. Everyone is personalized – as a researcher I just want quick, dirty access.”*

*“From an executive level I’m not a big user of that.”*

It is difficult to recommend an extensive personalization scheme based on these comments. It is possible that as the Knowledge Base evolves, with content growing and functionality added, consortium members may desire personalization to filter their view of the Knowledge Base. We see this as part of the expanded vision and not something to account for in the initial deployment.

However, even at the outset of the Initial Vision, a login and basic permissions scheme will be required. Broadly speaking, there are three approaches:

- Password-protect the whole Knowledge Base as a unit, i.e. Apache web-server “.htaccess” files;
- Role-based permissions such as “public,” “consortium member” or “I3P staff;”
- Item-based permissions, where each item or object is available or not to each individual login ID.

Once an application requires more than one role, any infrastructure built will support an unlimited number of roles. This is probably the best approach at the outset. Eventually, it is possible to imagine that the Core Vision, and certainly the Expanded Vision, would require item-based permissions, where individual access was tightly controlled. Note that this access control is not expected to be highly secure, i.e. for sensitive or confidential material, but simply to provide a view of the Knowledge Base customized for each type of user. Also note that we are suggesting that the users not be provided with the complexity to change their own interfaces, but simply that the I3P might eventually define a small set of views that are useful for, eg., “the public,” “consortium members” and “I3P staff.”

#### TEAM COLLABORATION

A team workspace is the one area where members feel the I3P could provide an infrastructure service for working groups.

*“If I have a large document it is easier for me to post it to a website, and tell the group to grab it from here. A lot of that will depend on what the working group is doing. One thing you might do is, when you create a work group, create a little space for the charter and working documents and so forth.”*

Extending this, allowing comments on documents received favorable response, but it seemed optional or peripheral to the needs. It was, however, the one “interactive” area that received any positive response at all. For instance, moderated discussion areas were universally discounted as useful for real work.

*“My own experience is that they [discussion forums] are of minimal value. You might offer to host it if people want, but my guess is you are not going to get many takers.”*

*“There is a really interesting, cultural phenomenon that goes on. You see people join these things to be proponents of their own views only, and they are not interested in positive discussion that is going someplace. I’ve even seen vendors get on and talk about problems with other people’s solutions. That is the kind of stuff you really want to avoid, as soon as people see that they leave.”*

Our recommendation is to consider collaboration tools as part of the Core Vision, but to add those features only very deliberately as consortium members request them.

# Knowledge Base Actors and Goals

*I3P STAFF*  
*ACADEMIC RESEARCHERS*  
*GOVERNMENT POLICY ANALYSTS*  
*INDUSTRY TECHNICAL LEADERS*  
*MEDIA JOURNALISTS*

## KNOWLEDGE BASE ACTORS AND GOALS

Actors are people who are attempting to meet some goal using the Knowledge Base. Note that actors are unrelated to the roles of the individual. A person with a role of CEO may be an actor who has a goal of searching for a white paper online, and this person may also be an actor who has a goal of updating their research profile in an expertise database.

The I3P staff and academic researchers will be the primary actors using the Knowledge Base. Over time other actors including government agency personnel, industry representatives and the general public may become active. For the time being, we concentrate our efforts on the day-to-day work of the staff and the week-to-week or month-to-month activities of the research community. We have also considered the media journalist because it is a straightforward scope that has potentially high value to the research community.

### *I3P STAFF*

The most frequent users of the I3P Knowledge Base will be the I3P staff. Initially, staff will add and update most content, while other actors will simply access content.

#### *General administrative functions*

- Create new user
- Edit user profile
- Delete user
- Reset password
- Create new webspace for working group
- Set user permission levels
- Set item permission levels
- Add help files
- Edit help files
- Load new design templates
- Review access logs
- Temporarily suspend single user access (account problem)
- Temporarily suspend system access (system upgrade)

#### *Expertise directory*

- Create new profile
- Edit profile
- Delete profile
- Edit meta information
- Add to search index

### *Funding directory*

- Create new agency entry
- Edit agency entry
- Delete agency entry
- Create new project entry
- Edit project entry
- Delete project entry
- Gather relevant funding opportunities
- Edit meta information
- Add record to search index

### *Topical news*

- Gather potential news items to queue
- Select from queue which news items to publish
- Add link to original news item
- Add abstract of original news item
- Edit meta information
- Archive the selected news items
- Bulk email the selected news items

### *Events calendar*

- Add event
- Edit event
- Delete event (rare)
- Archive event
- Add event meta information
- Add new calendar filter criteria

### *Digital Archives*

- Create meta-records
- Review meta-records
- Edit meta-records
- Add keywords
- Review keywords
- Edit keywords
- Add terms to search index
- Edit search index terms

### *ACADEMIC RESEARCHERS*

- Register username and password
- Confirm registration (email with web link)
- Update member profile

- Retrieve a lost username and/or password
- Subscribe to one or more email newsletters
- Suspend subscription to one or more email newsletters (vacation)
- Unsubscribe to one or more email newsletters
- Browse for existing related work
- Search for specific paper reference
- Search for colleagues with topical expertise
- Review calendar of upcoming conference deadlines
- Review funded projects
- Add a new project to the funding directory
- Post a paper to a working group webspace (or email the paper to the group liaison to post it)

*GOVERNMENT POLICY ANALYSTS AND  
INDUSTRY TECHNICAL LEADERS*

- Subscribe to one or more email newsletters
- Suspend subscription to one or more email newsletters (vacation)
- Unsubscribe to one or more email newsletters
- Review current work in the field
- Find researchers working on specific topics
- Find institutions with specific technical capability
- Find upcoming events to meet with researchers

*MEDIA JOURNALISTS*

- Read press releases
- Read background information
- Search for topical information and pointers
- Find expert on current news event
- Request interview with expert via I3P liaison

Capabilities  
Roadmap

## CAPABILITIES ROADMAP

We should remember that the “customers” for the I3P Knowledge Base are some of the most senior computer security researchers in the country. That is to say, they are smart, they are deeply immersed in computer technology, and they are at a point in their careers where they have seen all the likely errors before. It is important for the I3P to “first, do no harm.” Errors in judgment will do lasting damage to the I3P reputation. As one participant put it:

*“As we all find in our professional lives, it is not just the players, but the depth of the knowledge that comes with experience, and knowing where the pitfalls are, knowing the human dynamic, knowing how to build the team, knowing to know how to protect the budget. I think the team is going to have that kind of maturity. If you are going to grow the organization you have to have people with that depth of experience.”*

At this time it appears that the critical skill sets for success of the I3P Knowledge Base are experience in knowledge management and experience in software engineering. Patricia Erwin is now the I3P Knowledge Base Architect and Manager, and Jonas Meyer is in place as a software developer, satisfying both of these requirements.

In the realm of software development, not all software engineers are skilled software developers. There is a significant difference in mindset and skills between, say, a security research engineer, a network systems engineer, a database administrator, a data center manager and a software applications developer. Our recommendation for the I3P is that engineering job descriptions focus on software development skills and experience with integrated, normalized data models. The key skills are database modeling, page-flow design, clean architectural software design to allow for rapid evolution, and a wholistic systems worldview that accounts for end-users who are possibly much less skilled in information search and retrieval.

There are many organizational reporting structures possible, depending on the personalities and experiences of the team. The project managers and software developers will need a strong collaborative mindset and encourage contributions from the many interested constituents. They will frequently need to respond to suggestions for features unimagined by information architects and designers.

We do not expect systems hardware to be unusually expensive – the existing servers will be more than sufficient for the foreseeable future. Choosing to build vs. buy the software platform is an important strategic decision that should be determined after defining the initial and medium-term project. For the scope of this project, we recommend that an integrated data model be built from the start, even if this slows initial product deliverables. The advantages of starting with a good application design framework and then building the infrastructure for smooth product evolution outweigh the pressures to ship something immediately.

Appendix A:  
Synchronous Activities

## APPENDIX A: SYNCHRONOUS ACTIVITIES

While the focus of this work was on asynchronous collaboration tools, most people had something to say regarding meetings and in-person (synchronous) activities. When asked if in-person meetings were valuable, one typical response was:

*“Yes. Assuming there is something to do. For example, if it is just ‘let’s get together and talk,’ I would say no. But, if, on the other hand, it’s ‘here is this problem, let’s try to define the network problem, what exactly are we trying to do and here are the parameters’ – something like that might be very helpful, because you would have a lot of people trying to define what exactly the direction should be and how to get there.”*

Two people also mentioned that they go to other conferences already, and if the I3P could tie into those then it could decrease their travel time. A follow-on idea was that perhaps the I3P could work with other conference organizers to hold a panel discussion, or an informal “birds of a feather” gathering or other integrated conference activities.

*“Also, just sessions at conferences, pushing people to have a panel, or even having an I3P statement or panel, at say the conference on Computer Security.”*

Finally, several people mentioned videoconferencing and teleconferencing.

*“I teleconference quite a bit; I haven’t really done much video conferencing. I’ve always felt that the phone works just fine, I’ve never felt like I was missing something by not having the video, but of course if I used it extensively I might change my mind.”*

*“For 15 years I worked in an organization where I had people reporting to me from four different sites across the country. We played around with lots of different tools. The most effective things we found were a) forcing people to get together, holding off-sites and meetings and workshops and stuff like that. And b) video technology, low cost video technology, from the desktop, so you could see each other.”*

The I3P may be able to garner significantly increased participation of working groups between meetings by standardizing on some simple teleconference and/or videoconference protocols. Simply choosing a teleconference provider and arranging for them when required will prove to be good facilitation for working group process.

Appendix B:  
Research Participants

## APPENDIX B: RESEARCH PARTICIPANTS

Terry Benzel

University of California at Berkeley

Matt Bishop

Computer Security Research Lab at University of California at Davis

Chris Clifton

Purdue University, CERIAS

Ted Cooley

Thayer School of Engineering at Dartmouth College

Tracey Cote

Institute for Information Infrastructure Protection at Dartmouth College

Patricia Erwin

Institute for Information Infrastructure Protection at Dartmouth College

Adam Golodner

Institute for Security Technology Studies at Dartmouth College

Bob Gray

Institute for Security Technology Studies at Dartmouth College

John James

Dartmouth College Library

David Kotz

Institute for Security Technology Studies at Dartmouth College

Andrew Macpherson

Institute for Security Technology Studies at Dartmouth College

Susan McGrath

Institute for Security Technology Studies at Dartmouth College

Jennifer Merrill

Dartmouth College Library

Jonas Meyer

Institute for Information Infrastructure Protection at Dartmouth College

Ron Rivest

MIT Laboratory for Computer Science

Sujeet Sheno

Center for Information Security at the University of Tulsa

Gene Spafford

Purdue University, CERIAS

Paul Thompson

Institute for Security Technology Studies at Dartmouth College

Rae Zimmerman

Institute for Civil Infrastructure Systems at New York University

Marc Zissman

MIT Lincoln Laboratory

Additional input was gathered from the July 2003 Washington DC consortium meeting participants.