

# Personalized Information Environments

<http://www.cs.virginia.edu/~cyberia/PIE/>



James C. French, Edward K. O'Neil, Andrew Grimshaw  
{ french | oneil | grimshaw } @ cs.virginia.edu  
Department of Computer Science  
University of Virginia

Charles L. Viles  
viles@ils.unc.edu  
School of Information and Library Science  
University of North Carolina, Chapel Hill



## Overview

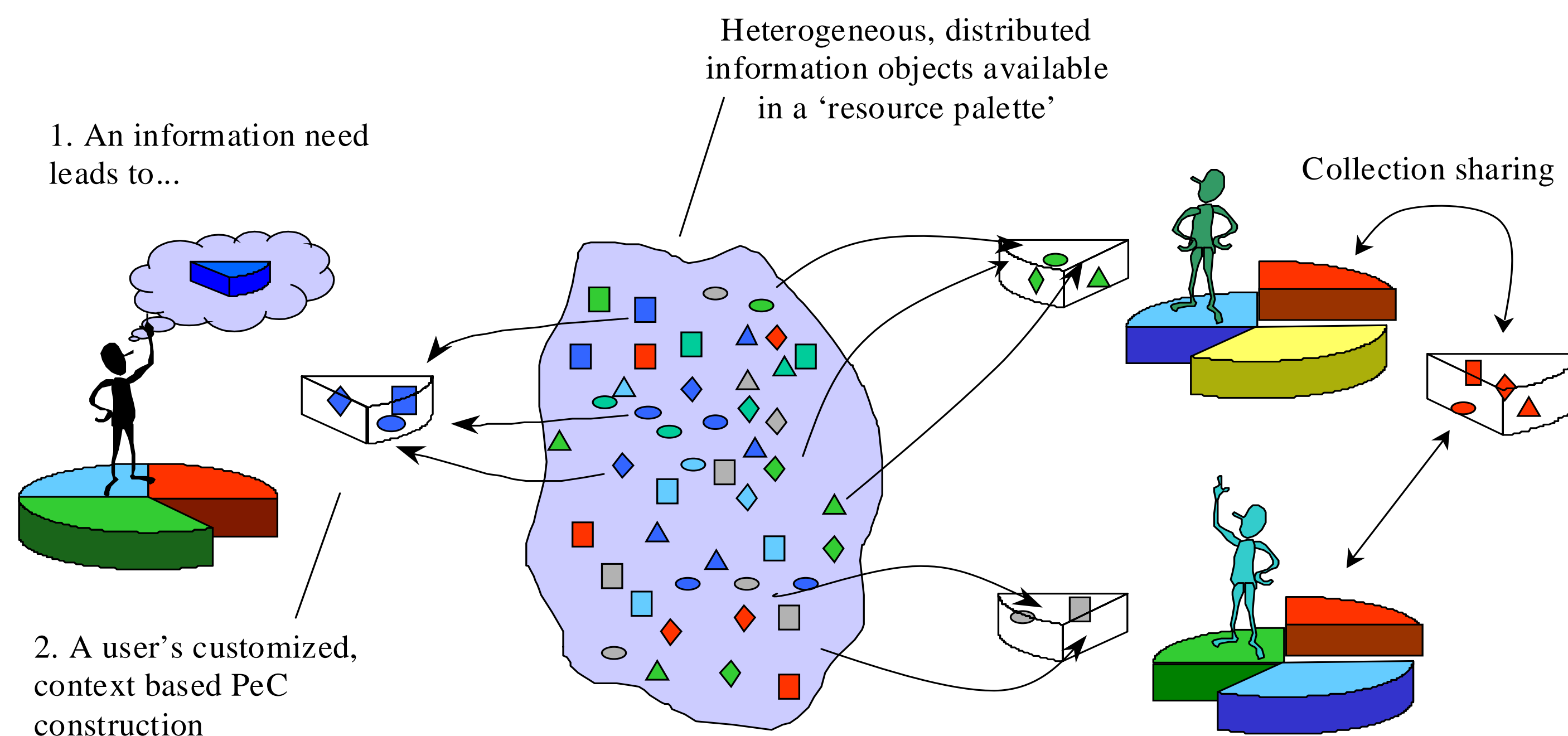
The Personalized Information Environment or *PIE* is a framework within which users may build and conduct highly customized searches on a distributed document collection of their own choosing. There are four driving principles behind the PIE.

- Customizability
- Efficient and Effective Search
- Controlled Sharability
- Privacy and Security

In contrast to a typical Internet search of multiple information resources, where control of which resources are searched is in the search engine's hands, a PIE places the control in the user's hands. In the PIE formulation, descriptions of resources are made available to users who decide which resources to include in a search. The process of resource selection is highly interactive and might involve sample searches and then selection or de-selection of resources from the user's current personalized collection. Regardless of the degree of interactivity, efficient and effective search is provided within whatever context the current collection of resources defines. Since a user may spend considerable effort customizing a personal resource collection, it makes sense to allow sharing of that collection in constrained ways or using pre-defined policies while maintaining whatever privacy or security constraints might be placed on particular resources or users.

## Collection Creation

In the PIE, collections are created by and for individuals or groups. Individuals maintain sets of preferences for their own PIE, and can manage as many Personal Collections (PeC) as they wish. An individual creates a new PeC by interactively selecting from a set of known resources. Once created, the PeC is persistent and is always available to that user until it is deleted.



## Prototype Implementation

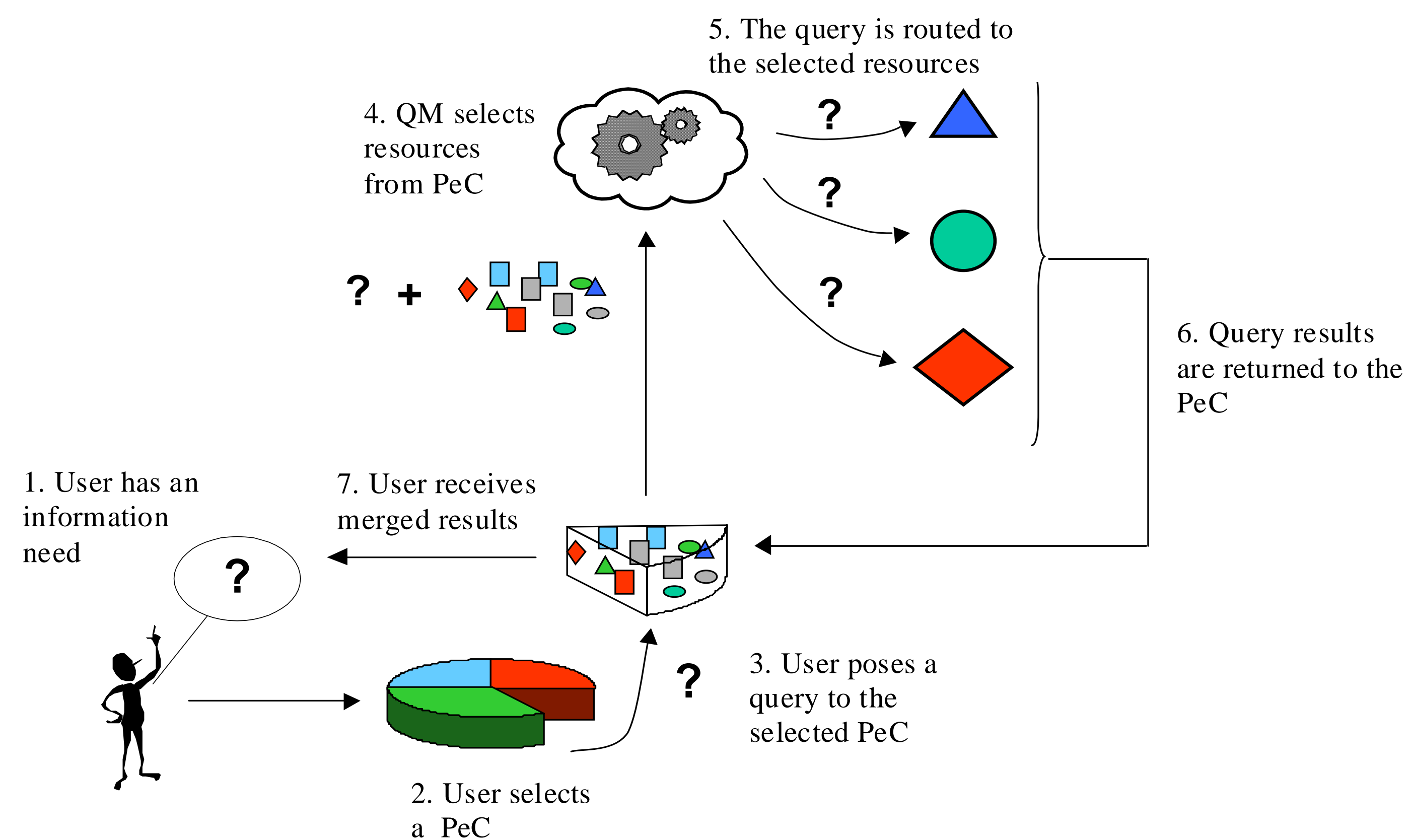
- 100+ sites from the NCSTRL collection (containing 30,000 technical reports)
- Document and Site level Dublin-Core compliant metadata crosswalks
- Interactive, persistent creation and use of personal collections
- Single machine, address-space disjoint Legion objects
- Registration with Query Mediator (QM) when Virtual Repositories are created
- Collection selection in QM using document frequency statistics

## Key Components

Component	Definition	Functionality
Virtual Repository (VIRP) 	A generic wrapper around a document repository.	Indexing, search, retrieval, metadata.
Personal Collection (PeC) 	A set of Virtual Repositories (or their surrogates) tuned to a particular topic or information need.	Search, add/delete/update resources, query filtering.
Query Mediator (QM) 	Given a query, chooses the set of VIRPs to which the query will be sent.	Filter resources based on query, add/delete/archive VIRP descriptions, monitor filtering performance.

## Search

User-level search is mediated by the PeC, and search for a particular query always happens within the context of the PeC to which the query was posed. If the PeC is aware of a Query Mediator (QM), then the PeC will always pre-filter the list of destination VIRPs by first sending the query to the QM. The QM then recommends a reduced set of VIRPs to which to send the query. The PeC then sends the query to this reduced set and merges the search results as they become available. Once assembled, the results are then sent back to the user. Efficiencies are realized when the user creates the PeC and when the query is posed to the PeC.



## Additional Features

- Collection sharing
- Client platform independence
- Persistent user and group profiles
- Extendable architecture for easily wrapping additional resource types

## Future Work

- User profile and document security
- User relevance feedback for tuning a PeC to specific information needs
- Expansion of available resource types

## Supporting Technologies and Standards

Legion (<http://legion.virginia.edu/>) is wide area metasytem software with bindings to MPI, distributed C++, PVM, and CORBA. All persistent objects in the PIE implementation are Legion objects or contained in Legion objects.



Dublin Core metadata standard (<http://www.purl.org/dc/>). A 15 element metadata set designed for cataloging electronic resources. The PIE implementation includes Dublin Core records for both sites and documents.



Networked Computer Science Technical Report Library or NCSTRL (<http://www.ncstrl.org/>). Approximately 30,000 bibliographic records from this operational digital library. Data is uploaded into the VIRP implementation using well-known methods from NCSTRL's Dienst protocol.



Extensible Markup Language (<http://www.w3.org/XML/>) is used for data interchange between Java objects in the PIE client and the C++ based persistent Legion objects.



This work is supported by DARPA contract N66001-97-C-8542.