

**Centro de Investigación Científica y de  
Educación Superior de Ensenada**



**CONSTRUCCION DE UN COLABORATORIO  
PARA ESTUDIOS DEL FENOMENO DE  
EL NINO EN MEXICO**

**TESIS  
MAESTRIA EN CIENCIAS**

**RENE FRANCISCO NAVARRO HERNANDEZ**

Ensenada, Baja California, Mexico. Enero de 1999



CENTRO DE INVESTIGACIÓN CIENTÍFICA Y DE EDUCACIÓN  
SUPERIOR DE ENSENADA



CICESE

DIVISIÓN DE FÍSICA APLICADA

CIENCIAS DE LA COMPUTACIÓN

CONSTRUCCIÓN DE UN COLABORATORIO PARA ESTUDIOS DEL  
FENÓMENO DE EL NIÑO EN MÉXICO

TESIS

que para cubrir parcialmente los requisitos necesarios para obtener el grado de  
MAESTRO EN CIENCIAS presenta:

RENÉ FRANCISCO NAVARRO HERNÁNDEZ

Ensenada, Baja California, México. Enero de 1999

**RESUMEN** de la Tesis de **René Francisco Navarro Hernández**, presentada como requisito parcial, para la obtención del grado de **MAESTRO EN CIENCIAS en CIENCIAS DE LA COMPUTACIÓN**. Ensenada, Baja California, México. Enero, 1999.

## **CONSTRUCCIÓN DE UN COLABORATORIO PARA ESTUDIOS DEL FENÓMENO DE EL NIÑO EN MÉXICO**

Resumen aprobado por:

Se presenta el desarrollo e implementación de un colaboratorio para estudios del fenómeno El Niño. Tiene como finalidad promover la creación de una red coordinada de investigación la República Mexicana en torno a este fenómeno. El colaboratorio se basa en un ambiente de visualización de datos científicos. Consiste de un conjunto de herramientas independientes de plataformas de hardware y software, entre las cuales se encuentran un visualizador de datos, pizarrón electrónico, chat box, visualizador de documentos e interfaz para el acceso de datos remotos. Estas herramientas se encuentran disponibles a través de Internet para aquellos investigadores de distintos estados de la nación que requieran hacer uso de ellas. Los datos se recuperan de un banco de datos en donde se almacena información generada por los diferentes grupos de investigación. Se hicieron pruebas de uso del colaboratorio con varios investigadores. Los resultados de las pruebas son alentadores, ya que los participantes pudieron operar fácilmente a el sistema. El colaboratorio puede operar en distintas instituciones académicas del país.

Palabras clave: Bases de Datos Científicas, Colaboratorios, Colaboración Científica, CSCW, EL Niño, Sistemas Colaborativos, Trabajo en Grupo.

**ABSTRACT** of the Thesis of **René Francisco Navarro Hernández**, presented as partial requirement to obtain the **MASTER IN SCIENCES** grade in **COMPUTER SCIENCES**. Ensenada, Baja California, México. January, 1999.

## **CONSTRUCTION OF A COLLABORATORY TO STUDY EL NIÑO IN MEXICO**

### **ABSTRACT**

The design and implementation of a collaboratory to study El Niño in México is presented. The main goal is to promote the creation of a coordinated research network around this phenomenon in the Mexican republic. This collaboratory is based on a collaborative environment for scientific data visualization. It includes a platform independent set of tools for data visualization as electronic white-board, chat box, document browser and an interface for remote access data. These tools are available through the Internet to any researcher. The data is retrieved from a remote scientific database which contains information generated by different research groups. Usability test were performed. Results indicated that the system is user friendly. The collaboratory is accessible to any academic institution.

**Keywords:** Scientific Databases, Collaboratory, Scientific Collaboration, CSCW, Collaborative System, Groupware, El Niño.

## DEDICATORIA

A la persona que me dio vida, mi madre Lourdes Hernández Venegas. Por el amor con el que me ha colmado durante todos estos años y por hacerme un hombre de bien. Gracias.

A mis hermanos Juan Antonio, María Auxiliadora, Jesús Alejandro, Rosalinda, Hernan Cornelio, Oscar Ramón y María Lourdes. Por su apoyo incondicional.

A la memoria de mi padre Jesús Navarro Zermeño, de quien conservo recuerdos muy gratos.

## AGRADECIMIENTOS

A mi director de tesis Dr. José Gómez Valdés. Por sus consejos y tiempo dedicados para que este trabajo fuese posible.

A los miembros de mi comité por su valiosa colaboración Dr. Jesús Favela Vara y Dr. José Luis Medina Monroy por sus acertados comentarios y por haberme ayudado a concluir este trabajo.

Al M.C. Cesar Enrique Rose Gómez, extraordinario maestro y excelente amigo. Gracias por su apoyo incondicional.

A mis apreciables amigos, “Los Villa”, Paloma y Manuel por compartir conmigo parte de su dicha y felicidad.

A mis compañeros de la maestría. Los compas: Armando Ayala, Antonio García, Antonio Rodríguez, Armando Carreon, Alfredo Cristóbal, Carlos González, Guillermo Licea, Humberto Chavez, Juan Contreras, Juan Garcilazo, Manuel Zapata, Manuel Alba, Martín Gaynor, Octavio García, Orlando Villegas, Rafael Llamas, Roberto Nuñez, Sergio Infante,.  
Las chicas: Carmen Monroy, Diana Ruiz, Gabby Rivera, Idalia Mendoza, Lilly Gómez, Marcela Rodríguez, Mary Avendaño, Urania Ceseña, Virginia Tamayo.

A mis amigos Abel Monjaraz, Mario Ortiz, Francisco Pérez y Ricardo Molina. Gracias por estar siempre allí brindándome su apoyo y amistad.

A todos los miembros de la lista de correos de excompañeros del ITH y pen pals: Ada Valenzuela, Betty Fierro, Betty López, Dennisse, Edith Grijalva, Fernando Encinas, Francisco Morales, Heriberto Huerta, Jallu Valdés, Jesus Federico, Leonor Sandoval, Mario Estrada, Nadia Garibay, Sandra Ballesteros. Por hacer mas llevadera mi estancia fuera de casa.

Al Centro de Investigación Científica y de Educación superior por el apoyo brindado para la obtención del grado de maestría.

Al Consejo Nacional de Ciencia y Tecnología, por haber proveído los medios para hacer posible este trabajo.

## CONTENIDO

CAPÍTULO I. INTRODUCCIÓN .....	1
I.1. OBJETIVO DEL PROYECTO DE TESIS .....	6
I.2. METODOLOGÍA .....	6
CAPÍTULO II. COLABORACIÓN EN LA INVESTIGACIÓN CIENTÍFICA .....	9
II.1. EL ESTUDIO DEL FENÓMENO EL NIÑO .....	9
II.2. EL CONCEPTO DE COLABORATORIOS .....	10
II.3. TECNOLOGÍAS DE LA INFORMACIÓN Y SU APLICACIÓN EN LA COLABORACIÓN CIENTÍFICA .....	13
II.4. EJEMPLOS DE COLABORATORIOS EN OPERACIÓN .....	20
CAPÍTULO III. BASES DE DATOS CIENTÍFICAS .....	25
III.1. DATOS CIENTÍFICOS .....	25
III.2. INTERCAMBIO DE DATOS CIENTÍFICOS .....	26
III.3. BANCOS DE DATOS CIENTÍFICOS .....	29
III.4. TÉCNICAS DE ALMACENAMIENTO DE DATOS CIENTÍFICOS .....	31
III.5. EL FORMATO DE ARCHIVO HDF .....	35
III.6. TECNOLOGÍAS DE ACCESO REMOTO A BANCOS DE DATOS CIENTÍFICOS .....	38
CAPÍTULO IV. DISEÑO E IMPLEMENTACIÓN .....	42
IV.1. DESARROLLO DEL PRIMER PROTOTIPO .....	43
IV.1.1. Java y las librerías de HDF .....	43
IV.1.2. Descripción de la arquitectura del prototipo .....	44
IV.1.3. Interface de usuario .....	48
IV.2. DESARROLLO DEL SEGUNDO PROTOTIPO .....	50
IV.2.1. RMI .....	51
IV.2.2. Descripción de la arquitectura del segundo prototipo .....	53
IV.2.3. Interface de usuario del sistema cliente .....	61
IV.3. DESARROLLO DEL TERCER PROTOTIPO .....	62
IV.3.1. Caracterización del ambiente colaborativo de visualización .....	64
IV.3.2. Descripción del ambiente de visualización colaborativo .....	68
IV.3.3. Arquitectura del ambiente colaborativo de visualización .....	74
CAPÍTULO V. EVALUACIÓN DEL AMBIENTE COLABORATIVO DE VISUALIZACIÓN .....	78
V.1. ESCENARIO DE EJECUCIÓN DE LAS PRUEBAS .....	78
V.2. RESULTADOS OBTENIDOS DE LAS PRUEBAS .....	80
CAPÍTULO VI. CONCLUSIONES Y TRABAJO FUTURO .....	82
VI.1. APORTACIONES .....	82
VI.2. TRABAJO FUTURO .....	83
LITERATURA CITADA .....	85



## LISTA DE FIGURAS

Figura	Descripción	Página
1	Modelo de colaboración científica.	12
2	Matriz espacio/tiempo para la clasificación de los sistemas groupware.	15
3	Tecnologías de groupware y su aplicación en el proceso de colaboración científica.	17
4	Utilización de herramientas groupware.	20
5	Arquitectura de CORE2000.	21
6	Arquitectura del ambiente CIF.	22
7	Instalaciones y equipo remoto accesibles a través del SPARC.	23
8	Contenido del descriptor de datos en una archivo HDF.	37
9	Organización física de un archivo HDF.	37
10	Ventana principal de la aplicación Java HDF Viewer (JHV).	44
11	Arquitectura del primer prototipo.	45
12	Modelo de objetos del primer prototipo.	46
13	Diagrama de secuencia para la recuperación de objetos de datos en un archivo HDF.	47
14	Componente HDFTreeView, representación gráfica de la jerarquía de objetos de una archivo HDF.	48
15	Interface de usuario del primer prototipo.	50
16	Arquitectura en capas del sistema RMI.	52
17	Arquitectura del segundo prototipo.	54
18	Modelo de objetos del sistema RemoteHDFServer.	55
19	Diagrama de secuencia de la creación del sistema RemoteHDFServer.	56
20	Sistema de nombramiento de RMI.	57
21	Diagrama de secuencia de la recuperación de objetos de datos de un archivo HDF remoto.	58
22	Diagrama de secuencia de la recuperación de objetos Vdata de un archivo HDF.	60
23	Despliegue de una tabla de los registros Vdata de un archivo HDF.	60
24	Representación de la jerarquía de objetos de datos de un archivo HDF remoto.	61
25	Interface de usuario del sistema cliente.	62
26	Modelo de un ambiente de colaboración.	63
27	Modelo de una ambiente colaborativo de visualización.	64
28	Formato de caracterización del espacio de diseño del ambiente colaborativo de visualización.	67

## LISTA DE FIGURAS (continuación)

Figura	Descripción	Página
29	Interface de usuario del ambiente colaborativo de colaboración.	68
30	Chat para el intercambio de mensajes textuales.	69
31	Pizarrón compartido para realizar anotaciones gráficas.	70
32	Visualizador de documentos hipertexto.	71
33	Lista de participantes en una reunión.	73
34	Arquitectura del ambiente colaborativo de visualización.	74
35	Diagrama de objetos del ambiente colaborativo de visualización.	75
36	Diagrama de secuencia del proceso de registro de un nuevo participante.	77

## LISTA DE TABLAS

Tabla	Descripción	Página
I	Resumen de las características de los laboratorios CEMS, SMCAL y SPARC.	24
II	Sumario de tendencias tecnológicas que influyen en el intercambio de datos e información científica.	28
III	Resumen de características de los formatos para datos científicos CDF, HDF y netCDF.	35
IV	Ejemplos de mensajes definidos por el protocolo de comunicación.	76

# Construcción de un Colaboratorio para Estudios del Fenómeno El Niño en México

## CAPÍTULO I. INTRODUCCIÓN

A partir de 1940, los avances de la computación y sus efectos en la ciencia han sido notables y benéficos. La computación permite analizar y resolver problemas cuya complejidad los hace imposibles de tratar de manera analítica, admite examinar y procesar grandes cantidades de datos, controlar instrumentos complejos, y facilita la comunicación. La colaboración, por otra parte, es elemento fundamental de la ciencia desde tiempos remotos. No obstante que las ciencias han sido muy beneficiadas por la computación en las últimas décadas, su impacto en el proceso de colaboración ha sido por ahora pequeño. Actualmente la colaboración científica se basa en interacciones cara a cara, reuniones de grupos, congresos, acciones individuales y experimentación en laboratorios [Kouzes, 96].

En los últimos años ha surgido un gran interés en la aplicación de la Tecnología de Información (TI) a la actividad científica. La reducción en la razón costo/rendimiento del poder de los equipos de cómputo y de los sistemas de telecomunicación, y la fusión de éstas dos tecnologías en una infraestructura que facilita la adquisición de información y la coordinación de esfuerzos, permiten desarrollar aplicaciones de software que van más allá de herramientas tradicionales del modelado y el análisis de resultados utilizadas comúnmente por los investigadores [Malone y Rockart, 1991]. Como ejemplo, podemos citar el surgimiento de *comunidades electrónicas* de investigadores que exploran nuevas formas de colaboración [BioMOO, 1997 y Chen *et al.*, 1996]. En este sentido, en los inicios de 1990, en un reporte del National Research Council de EUA se discute el concepto de

*colaboratorio* (collaboratory, palabra formada de la unión de los términos *collaboration* y *laboratory*). El término se define así: "un centro sin muros en el cual los investigadores de la nación, puedan realizar sus investigaciones sin importar su ubicación geográfica, interactuando con colegas, accediendo información, compartiendo datos y recursos computacionales; y accediendo información de bibliotecas digitales" [Cerf *et al.*, 1993]. Una de las finalidades de un colaboratorio es la de facilitar la creación de lazos entre los científicos de una área determinada, otra es la de promover la colaboración entre investigadores de diferentes disciplinas, así como acelerar el desarrollo y diseminación de conocimientos básicos, y minimizar el tiempo entre los descubrimientos y su aplicación. Se aprovecha la unión de la computación con la telecomunicación para aumentar el intercambio de información entre los investigadores [Kouses, 96]. Para esto, es necesario combinar los esfuerzos e intereses de la comunidad científica con los de la comunidad de la ciencias de la computación, con el fin de crear modelos de computación orientados a aplicaciones y sistemas de comunicación para apoyar la colaboración científica. Es aquí donde entra el *Trabajo Cooperativo Asistido por Computadora* (Computer Supported Cooperative Work), que busca ser una disciplina para encauzar el diseño directo y apropiado de software multiusuario y que proporciona soporte a las actividades en grupo (*groupware*). Groupware se define como la TI utilizada para ayudar a las personas a trabajar en equipo de manera más eficaz [Baecker, 1993].

Por otra parte, El Niño es un fenómeno físico en el sistema océano-atmósfera del Pacífico Ecuatorial el cual influye de manera significativa en la variabilidad climática a escala global. Entre estos cambios podemos citar el incremento en las precipitaciones pluviales en

diferentes regiones (i.e. Norteamérica y el Perú); sequías en regiones tales como el sudoeste de Africa, Australia, Filipinas e Indonesia entre otras más. Durante El Niño se altera la presión atmosférica en zonas muy distantes entre sí, se producen cambios en la dirección y la velocidad del viento y las zonas de lluvia de la región tropical se desplazan hacia latitudes altas. En el océano, la contracorriente ecuatorial, que desplaza las aguas frías de la corriente del Perú hacia el Oeste, se debilita, con lo cual se favorece el transporte de aguas cálidas hacia la costa de América del Sur [IAI, 1997].

El conocimiento del fenómeno El Niño ha mejorado substancialmente en la última década. Las nuevas técnicas para monitorear el océano y la atmósfera y los avances en el modelado dinámico, han permitido entender mejor cómo el sistema océano-atmósfera produce la variabilidad climática relacionada con el fenómeno. Esta percepción ha permitido el pronóstico de los periodos cálidos y fríos hasta con un año de anticipación. Para lograr los resultados anteriores se requiere de un gran volumen de datos para construir modelos precisos con los que sea posible explicar y predecir estos eventos. Además de conocer las causas del fenómeno, es de igual importancia conocer cuales son las repercusiones adversas y benéficas que están asociadas al mismo. Los beneficios resultantes de lo anterior consisten en la prevención y disminución de desastres humanos y económicos, puesto que mediante la utilización de información actualizada sobre la variabilidad climática y sus efectos, los gobiernos pueden prevenir a la ciudadanía de estos cambios y así los distintos sectores económicos se pueden preparar para afrontar alteraciones en los patrones de precipitaciones y temperaturas.

Los cambios en la temperatura de la superficie del mar influyen en los procesos que generan cambios en la salinidad de las aguas, alterándose por tanto, las condiciones ambientales de los ecosistemas marinos. Estos cambios, a su vez, afectan las poblaciones de peces, especialmente en las áreas del Pacífico Americano y, por ende, la actividad pesquera allí. Los cambios en la circulación atmosférica alteran el clima global, con lo que se afectan los recursos hidráulicos, la agricultura, la explotación forestal y otras actividades económicas importantes en extensas áreas del planeta. Además, de la variabilidad climática relacionada, se ha vinculado a este fenómeno la reaparición de enfermedades infecciosas en las regiones azotadas por lluvias más torrenciales que lo normal. Los científicos también investigan la creciente incidencia de "mareas rojas" tóxicas y otros fenómenos oceánicos y su posible conexión con temperaturas más altas [IAI, 1998].

Debido a la magnitud del impacto socioeconómico en las diferentes regiones afectadas por el fenómeno El Niño, en las últimas décadas se ha dado cada vez mayor importancia a su estudio por parte de grupos interdisciplinarios de investigadores en diferentes partes del mundo. Los esfuerzos de estos grupos están principalmente enfocados a [IAI, 94]:

1. Mejorar el entendimiento del fenómeno El Niño y las variaciones climáticas asociadas a él.
2. Mejorar los pronósticos climáticos.
3. Maximizar la utilidad de esta información para las regiones afectadas.

Lo anterior es una tarea compleja y laboriosa, y más para grupos de investigación aislados, debido a que resulta demasiado difícil y costoso para cualquiera de estos grupos

reunir, almacenar y analizar de manera separada, toda la información necesaria para entender completamente la variabilidad climática. Aun contando con los recursos suficientes, la misma complejidad inherente a estos tipos de problemas sobrepasa la capacidad intelectual y operacional de cualquier grupo. En México, hay varios grupos de investigadores dedicados al estudio del fenómeno de El Niño en distintas instituciones distribuidas a lo largo de todo el país. Las áreas de investigación son diversas, abarcando a la geofísica, oceanografía, biología y ecología entre otras. En diciembre de 1997 el Consejo Nacional de Ciencia y Tecnología (CONACyT) lanzó una convocatoria a la comunidad científica y tecnológica del país a concursar por financiamientos para proyectos cuyos esfuerzos estuvieran encaminados a documentar las causas, mecanismos y efectos del fenómeno de El Niño en la República Mexicana. En este concurso resultaron beneficiados un total de 52 proyectos de 18 instituciones de investigación y de educación superior ubicadas en 10 estados de la república, en donde se llevarán a cabo las investigaciones [CONACyT, 1998].

Dadas las repercusiones de este fenómeno de escala global, la comunidad científica mexicana tiene la tarea de estudiar las condiciones asociadas a él en la república, para poder contrarrestar los efectos dañinos y conocer las consecuencias benéficas. Para alcanzar el máximo aprovechamiento de los recursos humanos y materiales asignados a esta tarea, surge la necesidad de crear vínculos para: a) la cooperación entre los centros de investigación y universidades, b) el aprovechamiento de la infraestructura existente y de la experiencia acumulada, y c) el acceso a experimentos de campo, colecciones de datos y resultados entre los investigadores. El problema surge de algunos elementos que dificultan la colaboración, como por ejemplo, problemas de comunicación entre investigadores de disciplinas

diferentes, o entre instituciones con diferentes niveles de desarrollo tecnológico e inclusive la misma separación geográfica de las instituciones [IAI, 1997].

De lo anterior surge la motivación de este trabajo. El cual consiste en el diseño e implementación de un ambiente distribuido para la visualización colaborativa de datos, con la finalidad de ser puesto a disposición de los investigadores mexicanos involucrados en proyectos relacionados con el fenómeno de El Niño. El ambiente se basa en un conjunto de herramientas de software para la visualización de datos y de apoyo al trabajo en grupo accesibles a través de Internet proporcionando la infraestructura de comunicación requerida para apoyar la interacción entre los colaboradores y el medio de enlace para unificar los esfuerzos de distintos grupos de investigación del fenómeno en México.

### **I.1. Objetivo del proyecto de tesis.**

El objetivo de este proyecto es el diseño e implementación de un ambiente distribuido para apoyo a la colaboración científica a distancia entre los investigadores del fenómeno de El Niño en la República Mexicana.

### **I.2. Metodología.**

El ambiente a desarrollar tiene la finalidad de apoyar la interacción de los investigadores en el proceso de visualización y revisión de datos científicos (i.e. series de datos, imágenes, documentos) de manera distribuida. Este ambiente estará integrado por:

1. Interfaces de comunicación para el acceso compartido de información (i. e. series de datos, imágenes, documentos) almacenada en bancos de información remotos.
2. Herramientas que faciliten las distintas funciones de colaboración.

Como ejemplos de estas herramientas podemos mencionar las siguientes: *chat box* para el intercambio de mensajes textuales entre los participantes, un pizarrón de superficie compartida por los participantes en la que puedan hacer anotaciones por medio del ratón o algún otro dispositivo de señalamiento, acceso remoto compartido a documentos, entre otros.

Los pasos a seguir para el desarrollo de este sistema son los siguientes:

1. Identificar los requerimientos para la colaboración remota particulares al estudio del fenómeno El Niño.
2. Diseñar e implementar un sistema con arquitectura distribuida para el acceso remoto a información.
3. Diseñar e implementar las herramientas de software necesarias para apoyar la colaboración a distancia.
4. Realizar pruebas y obtener conclusiones sobre la facilidad de uso del ambiente.

En el desarrollo de sistemas groupware, es importante tener en consideración los elementos sociales involucrados en las actividades normales de un grupo de personas que interactúan con un mismo objetivo, que normalmente no se toman en cuenta en los sistemas tradicionales. Adicionalmente, es importante también considerar las cuestiones técnicas necesarias para la representación de las actividades asociadas a reuniones de trabajo en grupo, como son compartir objetos y recursos, y las implicaciones en las decisiones tácticas y estratégicas (i.e. sincronización, concurrencia, comunicación) en el diseño de estos sistemas [Roseman, 92]. En ese mismo sentido, con la finalidad de facilitar el diseño de sistemas colaborativos, Soriano [1996] creó una caracterización del espacio de diseño de éstos sistemas que permite al diseñador hacer referencia a esta representación, localizar los requerimientos y decidir sobre algunas propiedades del diseño. Con base en ello, Zapata [1997] define una metodología que considera las posibles relaciones entre los requerimientos funcionales y decisiones de diseño para ser usadas como reglas de diseño. Esta metodología será utilizada de manera auxiliar en la especificación de requerimientos y en el mapeo de éstos al diseño del sistema propuesto.

En lo referente al proceso global del desarrollo del sistema, seguiremos el paradigma de Orientación a Objetos enmarcado en el Modelo en Espiral [Boehm, 1988], que es un modelo evolutivo del ciclo de vida del desarrollo de software que une la naturaleza iterativa de un modelo de construcción de prototipos con los aspectos controlados y sistemáticos del modelo secuencial lineal.

## CAPÍTULO II. COLABORACIÓN EN LA INVESTIGACIÓN CIENTÍFICA

### II.1. El estudio del fenómeno El Niño.

Dado que la magnitud y complejidad del fenómeno El Niño lo hacen imposible de estudiar de una manera aislada, resulta necesaria la colaboración de grupos de investigadores de distintas disciplinas. En caso particular de nuestro país, son varias las instituciones involucradas en la investigación del fenómeno o en estudios de eventos relacionados al mismo. En su mayoría, estos proyectos son financiados con recursos otorgados por diferentes organismos gubernamentales tales como el Consejo Nacional de Ciencia y Tecnología (CONACyT), el cual en el presente año ha destinado fondos para la investigación del fenómeno por un monto de \$6,572,102.00 (seis millones quinientos setenta y dos mil ciento dos pesos mexicanos), con los cuales se financiaron 52 proyectos de 18 instituciones que se encuentran ubicadas en diez estados distintos de la república. Por razones económicas y estratégicas es de vital importancia el fomentar la cooperación entre estas instituciones en los procesos de investigación, recopilación de información y divulgación de los resultados obtenidos con la finalidad de maximizar el aprovechamiento de los recursos humanos y materiales que se tienen a la mano. Resulta así necesario la promoción del intercambio de datos científicos así como incrementar la capacidad de recolección y análisis de información a través de programas de cooperación mediante el establecimiento de una red de investigación coordinada, logrando de ésta manera una masa crítica de la infraestructura nacional que ayude a encauzar el estudio del fenómeno, hacia aplicaciones de relevancia social y económica para el país.

Las principales dificultades que plantea esta empresa son en esencia de carácter social y tecnológico. Para lograr la integración de las instituciones nacionales es necesario la creación de vínculos de colaboración. Inherentes al proceso de colaboración están asociadas distintas actividades como la coordinación, la comunicación, la planeación y la toma de decisiones. Estas actividades se ven afectadas, entre otras razones, por la separación geográfica de las instituciones, por problemas de comunicación entre investigadores de las diversas disciplinas que involucra, de manera necesaria, el estudio del fenómeno, así como también por la diferencia en los niveles de desarrollo tecnológico que se presentan en las instituciones.

## **II.2. El concepto de laboratorios.**

En 1993, Wulf [Cerf, 1993] encabezó un panel de investigación cuyas conclusiones fueron proponer una estrategia para incrementar la eficiencia de los recursos de investigación mediante la combinación de la computación y las tecnologías de comunicaciones electrónicas. El panel sugirió el diseño e implementación de sistemas integrados de comunicación y computación orientada hacia herramientas con las que se pudiera apoyar el trabajo en colaboración entre científicos de diferentes localizaciones geográficas. El término escogido para identificar este concepto fue *colaboratorio*. Definido en forma más concreta como la combinación de tecnologías, herramientas e infraestructura que permite a los científicos trabajar en instalaciones remotas e interactuar con otros investigadores como si estuvieran localizados en el mismo lugar [Greenberg, 1990].

Entre la principal razón que justifica la aplicación de la tecnología está la escasez o reducción de fondos para financiar proyectos de investigación científica. En esta situación, el reto para los investigadores es buscar la manera de evitar un decremento de la producción científica. Otra razón de peso es que en la medida en que el equipo o instrumentación se vuelve más complejo y costoso, resulta más atractivo el desarrollo de sistemas distribuidos para su control, así como el compartir datos entre los investigadores interesados [Agarwal *et al.*, 1998]. Aparte de estas razones prácticas y económicas, tenemos también que en el quehacer científico la colaboración es un factor inherente a la actividad. De alguna manera los laboratorios han servido como escenarios físicos para el establecimiento de organizaciones sociales que permiten a los investigadores interactuar unos con otros. Permiten también, mediante la utilización del equipo, certificar y diseminar el conocimiento, así como la capacitación de futuros investigadores [Kouses, 1997]. La investigación es un proceso social, en el que se requiere de mucha interacción por parte de los investigadores participantes para iniciar y dar seguimiento a una investigación que se realiza en colaboración. Para describir este proceso el siguiente modelo (Figura 1) explica las distintas fases de la colaboración en la investigación científica [Kraut *et al.*, 1988].

Nivel de relación	Encontrar un colega	Supervisar y mantener el progreso	Establecer la división del crédito
	Compartir historial de conjeturas	Establecer la división del trabajo	
	Identificar intereses compartidos	Entablar confianza	
Nivel de tarea	Generación de ideas y planeación	Compartir información	Escritura del artículo
		Coordinación actividades	
		Hacer el trabajo	
	Iniciación	Ejecución	Presentación Pública

Figura 1. Modelo de colaboración científica

Los tipos de colaboración científica abarcan un rango muy amplio en términos del tamaño del grupo, el estilo de colaboración, el enfoque (i.e. experimental, teórica, computacional) Kouses *et al.* [1996] identificaron cuatro categorías generales:

- *Igual a igual*: En este tipo de colaboración los investigadores comparten el mismo nivel jerárquico y un vocabulario científico común. Los aspectos más importantes de colaboración son el intercambio de ideas, compartir instrumentos, compartir datos crudos o ya analizados.
- *Mentor/Estudiante*: En algunos casos la colaboración involucra a investigadores de categoría más alta y sus asociados de menor rango, como estudiantes o visitantes de postdoctorado. En este caso el mentor puede hacer uso de material preparado con anterioridad para la instrucción de técnicas de adquisición de datos, técnicas de análisis y demostración de principios científicos.

- *Interdisciplinaria*: En algunas ocasiones la colaboración involucra a investigadores realizando estudios complementarios del mismo sistema o fenómeno. Los investigadores en este tipo de colaboración pueden no compartir un vocabulario común, por lo que frecuentemente traducen los resultados a sus propios términos. Así mismo los investigadores pueden alternarse los roles de mentor/estudiante cuando pretenden sintetizar sus descubrimientos.
- *Productor/Consumidor*: Algunos tipos de colaboración implican la interacción de investigadores de distintas disciplinas, pero en este caso un investigador o equipo de investigación sule de información a otro. Los colaboradores en este tipo de relación tienen pocas oportunidades para el contacto profesional. En general, le dan más importancia en recibir o entregar el trabajo a la otra parte.

Es importante hacer notar el hecho de que una misma colaboración puede involucrar a varios elementos de las distintas categorías de colaboración descritas, ya sea en forma simultánea o en alguna de las etapas de la investigación. Estos modelos ayudan a mostrar las diferentes necesidades de comunicación que tienen los investigadores a medida que actúan en distintos roles y cómo estas mismas necesidades podrían variar, según cambia la tarea o naturaleza de la colaboración.

### **II.3. Tecnologías de la información y su aplicación en la colaboración científica.**

Un colaboratorio facilita la interacción científica dentro de un grupo mediante la creación de un ambiente virtual en el cual los investigadores puedan interactuar. Para que

puedan cumplir con su finalidad, este ambiente debe de ser socialmente aceptado por las personas que participan, mejorando las habilidades de las personas para realizar tareas. Con esta finalidad es necesario construir las herramientas de software necesarias e integrarlas de tal manera que la interacción entre los usuarios sea de la forma más natural posible. En este tipo de aspectos es donde tiene su aplicación el trabajo cooperativo asistido por computadora o CSCW (Computer Supported Cooperative Work , por su siglas en inglés). Es un campo relativamente nuevo que involucra diferentes disciplinas en el estudio de las características de la comunicación, colaboración y coordinación entre grupos de trabajo, y la utilización de la computadora como elemento que facilita la realización de estos procesos [Grudin, 1991]. A los sistemas de software multiusuario que asisten a estas actividades se les denomina software para grupos o groupware. Una definición más específica de los sistemas groupware es que son sistemas computarizados que apoyan a grupos de personas involucrados en una tarea (u objetivo) común y que proporcionan una interface hacia un ambiente compartido [Ellis *et al.*, 1991].

Podemos clasificar los sistemas groupware en cuatro categorías, considerando las nociones de tiempo y espacio en las que se dan las interacciones que auxilian estos sistemas [Ellis *et al.*, 1991]. Con respecto al tiempo, los tipos de interacción se dividen en interacciones síncronas, cuando la interacción se da a un mismo instante o en tiempo real, o interacciones asíncronas, cuando la interacción se efectúa en tiempos diferentes. En lo que se refiere al espacio, la interacción se puede realizar al encontrarse las personas en un mismo lugar o cuando las personas se encuentran distribuidas en distintos lugares. Partiendo de

estas divisiones y formando las posibles combinaciones llegamos a cuatro categorías ilustradas en la Figura 2.

	Mismo Tiempo	Diferente Tiempo
Mismo Lugar	Interacción cara a cara	Interacción asíncrona
Diferente Lugar	Interacción distribuida sincrónica	Interacción distribuida asíncrona

Figura 2. Matriz espacio/tiempo para la clasificación de los sistemas groupware.

Los ejemplos de sistemas de groupware son múltiples: el correo electrónico, sistemas de teleconferencia por computadora, boletines electrónicos, sistemas para la edición/revisión colaborativa de documentos, sistemas de apoyo a la toma de decisiones, sistemas para la administración de flujos de trabajo “workflow systems”, salas de reuniones computarizadas, sistemas de interacción casual, salones virtuales, entre otros. Aplicadas a la investigación, estas herramientas facilitan la interacción de colegas remotos: compartiendo datos y recursos, revisando instrumentos distantes y equipo de control, administrando el proyecto de investigación, y participando en la toma de decisiones.

Lederberg [1989] describe un modelo del ciclo de vida del proceso de investigación científica enumerando las distintas etapas que involucra este proceso, las cuales son:

1. Encontrar la motivación y percepción social de un área.
2. Construir un marco teórico y formular preguntas.
3. Diseñar un enfoque experimental para resolver estas preguntas.
4. Conseguir los recursos incluyendo el espacio, las personas, el equipo y el dinero.
5. Dirigir experimentos, posiblemente utilizando técnicas de administración de proyectos.
6. Analizar datos.
7. Formular y comprobar conclusiones y teorías.
8. Presentar resultados preliminares a colegas a través de borradores de manuscritos y presentaciones informales.
9. Publicar y presentar formalmente los resultados en revistas, conferencias, etc.
10. Recibir críticas de la comunidad.
11. Atraer nuevos participantes, establecer una nueva organización, extender el proyecto.

Greenberg [1990] extiende este modelo añadiendo el mapeo de las funciones hacia las herramientas de groupware que las auxilian y a las tecnologías que fundamentan estas herramientas (Figuras 3a y 3b).

<i>Funciones de investigación y desarrollo</i>	Educación, socializar, motivar al investigador	Construir un marco teórico, formular preguntas	Diseño de la estrategia	Agrupación de recursos (espacio, personas, equipo, dinero)	Administración y organización del proyecto	
<i>Funciones de colaboración</i>		Generación de ideas	Bases de conocimiento colectiva (adquisición, asimilación y diseminación )	Búsqueda colaborativa de recursos Encontrar un colaborador	Administración y organización del proyecto Capturar, desplegar editar el estado de colaboración Negociación y asignación de recursos escasos Apoyo a toma de decisiones	Experimentos colaborativos remotos Telepresencia Construcción rápida de prototipos y transporte de artefactos
<i>Herramientas de colaboración</i>	Entrenamiento de usuarios, herramientas de educación	Bibliotecas digitales Estándares de acceso	Sistemas de recuperación de información	Discusión estructurada	Sistemas de apoyo a toma de decisiones en grupo	Experimentos remotos y planificación de recursos
<i>Tecnología habilitadora</i>	Ingeniería de software Tecnología para el diseño de interfaces multiusuario Herramientas de ingeniería de software	Creación tecnologías y herramientas requeridas Instrumentación de herramientas para evaluar su impacto	Estándares (gráficos, ventanas, sistemas, información)	<u>Gráficos, interfaz humano-computadora &amp; multimedia</u> Ventanas y gráficos virtuales a través de diferentes máquinas Ventanas y espacios compartidos		Voz/sonido Vídeo Hardware para despliegue Despliegues públicos Dispositivos de despliegue y sistemas de entrada transportables

Figura 3a. Tecnologías de groupware y su aplicación en el proceso de colaboración científica.

<i>Funciones de investigación y desarrollo</i>	Analizar datos	Plantear y probar conclusiones	Presentación a colegas	Publicar	Obtener reacción de la comunidad	Comercializar	Atraer nuevos participantes/formar nuevas organizaciones
<i>Funciones de colaboración</i>	Manipular el ambiente físico		Revisión y comentarios en forma abierta	Preparar documentos Publicar documentos Revisar Documentos	Seminarios y reuniones distribuidas Videoconferencia Comunicación síncrona y asíncrona Espacio de gesticulación compartido	<u>Reuniones</u> Colaboración informal Administración en tiempo real de la reunión Memoria organizacional (tomar notas)	
<i>Herramientas de colaboración</i>	Boletines electrónicos Conferencia asíncrona	Herramientas para coautoría Superficies y herramientas para dibujar compartidas	<u>Reuniones v Edición remota</u> Salas de reuniones electrónicas		Vídeo teléfonos y salones de conferencia Salones virtuales	Coordinación de conferencia Protocolos para el paso de cursor y controles para la edición en grupo	
<i>Tecnología habilitadora</i>	Robótica Tele-operación	Redes con alto ancho de banda Comunicación móvil Servicios de enlace	<u>Comunicación</u> Transmisión de vídeo Herramientas para el desempeño de la red, depuración y administración		<u>Recuperación e indexado de Información</u> Representación de documentos, edición, acceso Búsquedas lingüísticas rápidas Extracción y segmentación de características imágenes de manera automática		Bancos digitales de imágenes Hipertexto & hipermedios Bases de datos compartidas

Figura 3b. Tecnologías de groupware y su aplicación en el proceso de colaboración científica (cont.).

Actualmente hemos identificados una decena de laboratorios trabajando en forma operacional o como prototipos. Las herramientas comúnmente utilizadas en estas implementaciones incluyen:

- *Instrumentación en línea, cómputo y visualización*: Sistemas para la adquisición, análisis, procesamiento y visualización en forma distribuida.
- *Interfaces de comunicación*: Permiten el acceso remoto a bases de datos científicas, repositorios de información documental y bibliotecas digitales.

- *Audio/video conferencia*: Los investigadores pueden interactuar entre ellos, monitorear instrumentos e instalaciones.
- *Chat box*: Con esta herramienta los colaboradores pueden intercambiar mensajes textuales.
- *Pizarrones electrónicos y ventanas compartidas*: Con estas aplicaciones los colaboradores pueden interactuar manipulando objetos desplegados en las ventanas compartidas y realizar anotaciones sobre imágenes o gráficas desplegadas.
- *Cuadernos electrónicos*: Proveen acceso distribuido a datos, así como captura automatizada de datos, búsquedas, y otras funciones que no son posibles de realizar en los cuadernos tradicionales.

La Figura 4 muestra la utilización de estas herramientas en distintas actividades de colaboración y cómo es que estas herramientas proporcionan distinta funcionalidad acorde a la naturaleza del intercambio de información, ya sea dinámico o estático, de las interacciones, y de la naturaleza de la sesión, ya sea síncrona o asíncrona. Por ejemplo, en el caso de teletutoría y teleconferencia la interacción se lleva a cabo en tiempo real. Sin embargo, existen diferencias en la forma de intercambiar información. En el caso de la teletutoría generalmente el tipo de colaboración es mentor/estudiante, por lo que el flujo de información se da principalmente en un sólo sentido: del mentor al estudiante. Por otra parte, las sesiones de teleconferencias involucran un intercambio de información más activo por parte de todos los participantes de la sesión.

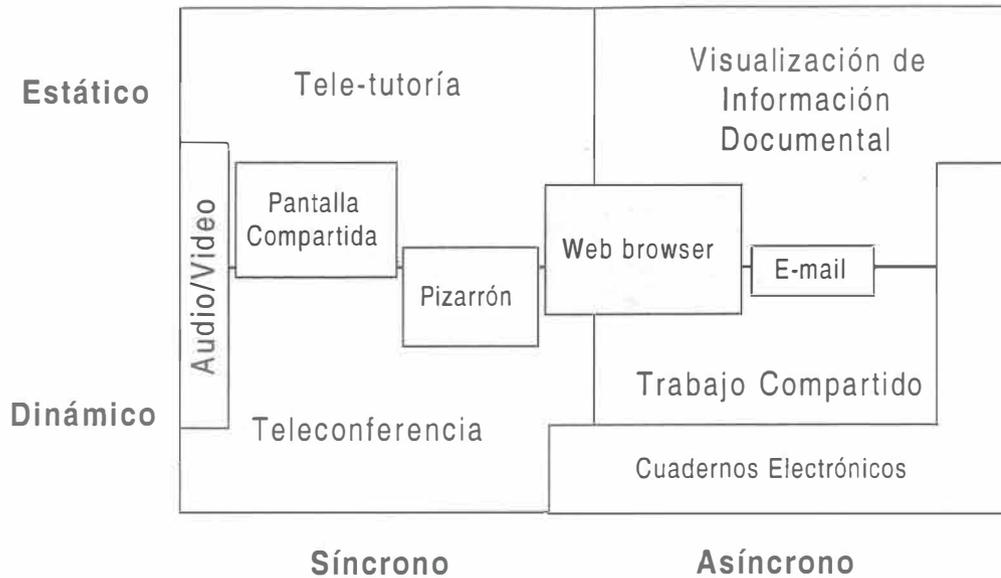


Figura 4 Utilización de herramientas de groupware

#### II.4. Ejemplos de laboratorios en operación.

A continuación se presenta una breve reseña de tres laboratorios actualmente en operación. El primero es el Collaboratory for Environmental Molecular Sciences (CEMS) el cual es una extensión del Environmental Molecular Sciences Laboratory (<http://www.emsl.pnl.gov:2080/docs/collab/>). Ambos se encuentran ubicados en el Pacific Northwest Laboratory en Richland, Washington, el cual está patrocinado por el U.S. Department of Energy (DOE). La misión de este laboratorio es reunir expertos de diferentes disciplinas, química y física principalmente, en una comunidad electrónica para colaborar en la investigación y en el desarrollo de tecnologías para la restauración y conservación del medio ambiente.

Los científicos ambientales utilizan técnicas experimentales muy variadas para entender los sistemas moleculares. Los instrumentos utilizados incluyen grandes espectroscopios de resonancia magnética nuclear y espectroscopios de masa, así como

dispositivos de pequeña escala para el estudio de la física y química molecular. El proyecto piloto con el que inicio a operar este colaboratorio fue la puesta en línea en 1996 del espectrómetro Laser Desorption RF Ion Trap Mass, para que éste pudiera ser controlado en forma remota y permitiera la adquisición de datos a través de Internet. Actualmente se encuentra en desarrollo un banco de datos científicos. Educadores de distintas universidades o colaboradores científicos pueden hacer uso del banco y otros instrumentos para la educación, monitoreo y adquisición de datos a distancia. El colaboratorio está basado en el sistema Collaborative Research Environment (CORE2000) el cual provee las facilidades para iniciar o incorporar sesiones de trabajo y utilizar varias herramientas. Este ambiente se desarrolló utilizando el marco de clases para construir aplicaciones colaborativas Habanero. La arquitectura de este ambiente se presenta en la Figura 5.

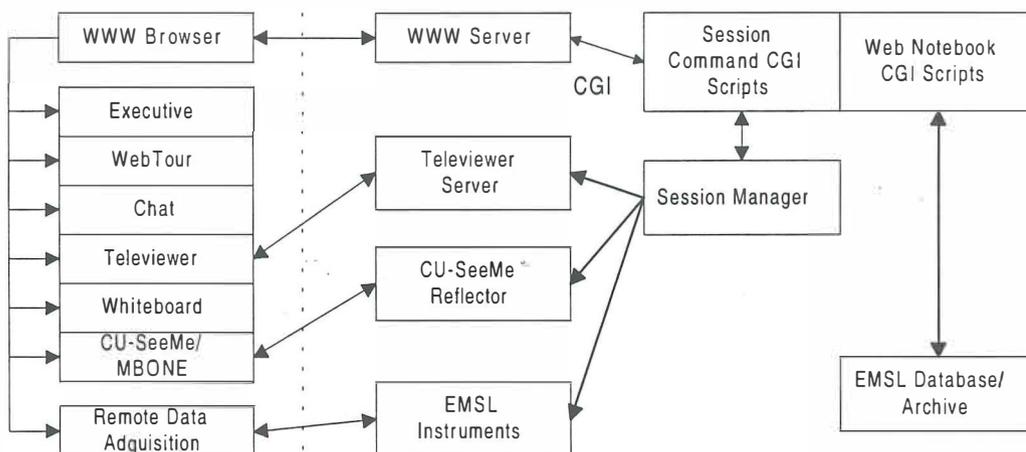


Figura 5. Arquitectura de CORE2000

El segundo colaboratorio de importancia es el Spectro-Microscopy Collaboratory (SMCAL) del Spectro-Microscopy Facility of the Advanced Light Source (ALS) Beamline 7.0 en el Lawrence Berkeley National Laboratory (<http://www-itg.lbl.gov/BL7Collab/>). Este colaboratorio permite el acceso a tres herramientas analíticas que proveen información

espacial sobre materiales en escalas de micras hasta átomos. Los colaboradores que utilizan estos instrumentos se encuentran distribuidos en nueve institutos de investigación.

El SMCAL se basa en el ambiente Collaboratory Infrastructure Framework e incluye distintas tecnologías de software (i.e. Java, CORBA, SSL, LDAP) para la interconexión de componentes y herramientas de software para el desarrollo de laboratorios del DOE (Figura 6). Estos servicios se proporcionan a través de una interfaz para la programación de aplicaciones para el lenguaje Java.

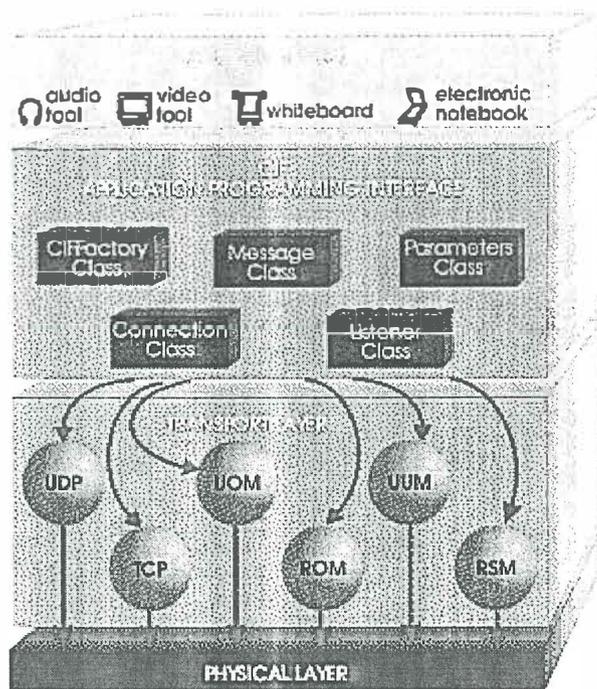


Figura 6. Arquitectura del ambiente CIF

El Space Physics and Aeronomy Research Collaboratory (SPARC) es el colaboratorio que actualmente presenta un mayor desarrollo (<http://www.crew.umich.edu/UARC/>). Puesto en marcha en 1992 con el nombre de Upper Atmospheric Research Collaboratory es el resultado del trabajo de distintos departamentos y laboratorios de la Universidad de

Michigan. En su desarrollo involucra a científicos de la atmósfera superior, física espacial, ciencias de la computación y ciencias de la conducta. En este laboratorio se realizan estudios de fenómenos que se presentan en la atmósfera superior tales como tormentas magnéticas originadas en el sol. Estas tormentas envían cantidades enormes de partículas cargadas hacia la tierra, lo que ocasiona trastornos e interferencias en la recepción de señales de radio y televisión, perturbaciones en la transmisión de energía eléctrica que ponen en peligro a naves espaciales en órbita. En este laboratorio los investigadores desde su oficina pueden controlar y recolectar datos de varios instrumentos localizados alrededor del planeta o en órbita espacial como se muestra en la Figura 7.

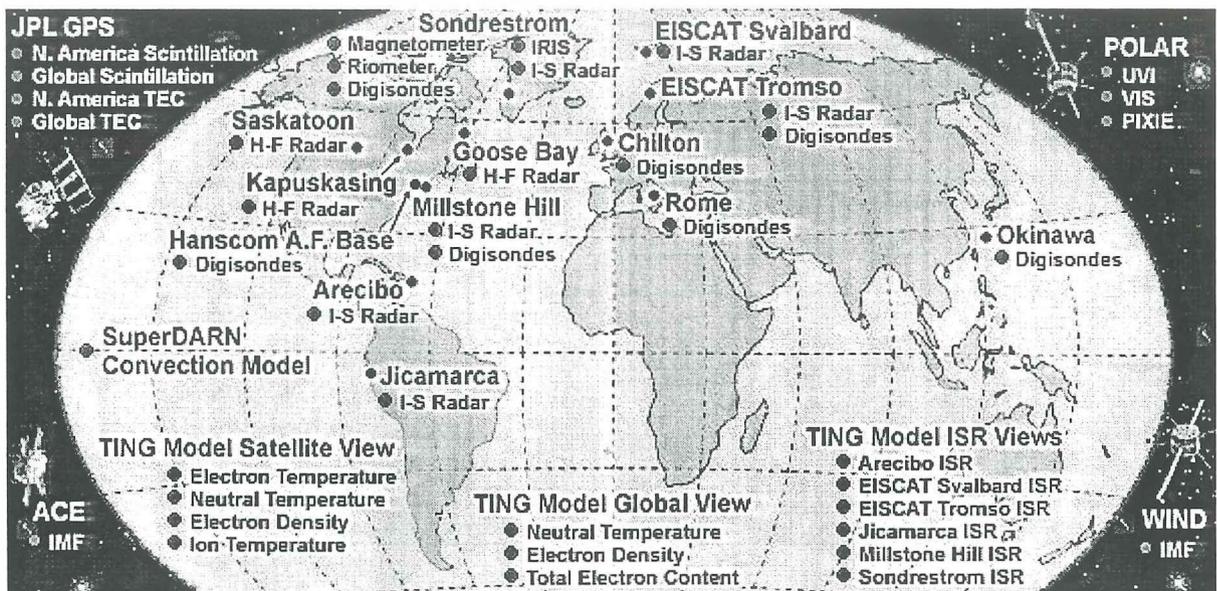


Figura 7. Instalaciones y equipo remoto accesible a través del SPARC.

El ambiente de colaboración utilizado en el SPARC es el Collaboratory Builder Environment (CBE). Originalmente desarrollado bajo la plataforma de desarrollo de NextStep, el CBE es un conjunto de “applets” de Java que incluyen visualizadores de

instrumentos, pizarrones electrónicos, “chat” integrados mediante un administrador de sesión.

La Tabla I presenta una comparación de las características de los tres laboratorios anteriormente descritos.

Tabla I. Resumen de las características de los laboratorios CEMS, SMCAL y SPARC.

	CEMS	SMCAL	SPARC
Area de Investigación:	Ciencias Ambientales, química y biología moleculares	Química molecular	Atmósfera superior y física del espacio
Experimentación Remota			
Instrumentos en Línea	<ul style="list-style-type: none"> <li>•Radio Frequency (RF)</li> <li>•Ion Trap Mass Spectrometer</li> </ul>	<ul style="list-style-type: none"> <li>•Spectro-Microscopy</li> </ul>	<ul style="list-style-type: none"> <li>•Incoherent scatter radar</li> <li>•Imaging riometer</li> <li>•Magnetometer</li> <li>•Optical spectrometers</li> <li>•Photometers</li> </ul>
Adquisición de Datos	✓	✓	✓
Base de Datos Compartida	✓	✓	✓
Herramientas			
Pizarrón Electrónico	✓	✓	✓
Chat	✓	✓	✓
Televiewer	✓	✓	✓
Videoconferencia	mbone, CU-SeeMe	mbone	N/D
Visualizador de Documentos	✓	✓	✓
Implementación	CORE 200(Java)	CIF	CBE

## CAPÍTULO III BASES DE DATOS CIENTÍFICAS

### III.1. Datos científicos.

Toda forma de investigación científica involucra la interacción formal e informal con otros investigadores y el intercambio de datos. La colaboración aumenta la capacidad de los investigadores para el avance del conocimiento. Los datos son los elementos con los que se construye el conocimiento. Con estos es posible la generación de nuevas teorías, conceptos y modelos. Son también el medio por el cual se transforman estos nuevos conceptos en aplicaciones para el beneficio de la sociedad. La recolección de datos es el paso inicial en el proceso de creación, organización y la aplicación del conocimiento [Westbrook,1992]. Este proceso implica el registro, análisis y diseminación de los datos. Existen diferentes maneras de caracterizar los datos científicos. Una de ellas es por su forma, ya sean numéricos, simbólicos, imágenes fijas, animaciones, o alguna otra. Otra es por la manera en la que fueron generados o recolectados, ya sea vía experimentos, observación o simulación, y otra por su nivel de calidad [NRC, 1997].

Existen también diferentes niveles jerárquicos para la clasificación de los datos de acuerdo a su refinamiento. En los niveles superiores tenemos a los datos experimentales y a los datos observacionales. En lo referente a los datos experimentales, en el primer nivel se encuentran sólo series de números o datos crudos. Normalmente publicados como tablas y su distribución es limitada. En un segundo nivel tenemos la publicación de resultados obtenidos mediante datos recolectados, en ocasiones se incluyen estos datos o se describe como se obtuvieron. En el siguiente nivel se encuentran aquellos datos que son el resultado

de una compilación de diferentes fuentes con la finalidad de encontrar los mejores valores para los datos registrados. Estos registros son ampliamente distribuidos y los encontramos en foros especializados.

De manera semejante, los datos de observaciones pueden ser procesados e interpretados a diferentes niveles de complejidad. Los datos originales sin procesar o mínimamente procesados son comúnmente los más difíciles de comprender o utilizar [NRC, 1995]. Cada nivel de procesamiento agrega cierto valor a los datos crudos, resumiéndolos, sintetizando nuevos productos o proveyendo una interpretación de los datos originales. Con esto, los datos se hacen más fáciles de entender y están mejor documentados para usuarios no expertos. Se espera que estos datos de más alto nivel sean los que tienen el más alto valor para su preservación e intercambio, ya que son más fáciles de entender y pueden ser utilizados por un mayor número de usuarios. Normalmente lo que ocurre es lo contrario, solo con los datos originales sin procesar es posible recrear los niveles superiores de procesamiento. Lo que justifica la conservación de los datos originales.

### **III.2. Intercambio de datos científicos.**

El uso particular de los datos y las características de las distintas disciplinas científicas determinan la necesidad y condiciones de acceso a bancos de datos en esas áreas de investigación. La observación de fenómenos naturales tales como las circulación atmosférica producen cantidades enormes de datos. La característica principal de este proceso es la gran cantidad de escalas espacio-temporales. Su comprensión requiere de observaciones con suficiente resolución espacial y temporal para describirlos [Eltgroth *et al.*, 1997]. La

circulación atmosférica controla el clima del planeta produciendo grandes variaciones a escalas de tiempo en el rango de horas a décadas y a escala espacial en un rango de menos de 1 Km hasta miles de kilómetros. El pronóstico del clima para más de un día requiere de la adquisición, procesado e interpretación de grandes volúmenes de observaciones a escala continental, por lo menos.

De la disponibilidad de estos datos, tanto como la de los productos de su procesamiento y los resultados de investigaciones experimentales, depende en gran medida el beneficio público que puedan generar los investigadores. El objetivo de todo investigador es que otros investigadores encuentren útil su trabajo y lo usen. Los investigadores, y la sociedad en general, resultan beneficiados en mayor medida con el intercambio libre de información.

El almacenamiento e intercambio de datos científicos han sido más problemáticos que su recolección. Aun en las mismas comunidades donde se generan y se utilizan los datos, su almacenamiento y diseminación requiere de un gran esfuerzo y tiene un alto costo económico. Por otra parte, muchos de estos datos tienen un valor significativo fuera de su lugar de origen, lo que hace aun más importante el ofrecer el acceso libre a ellos. Las nuevas tecnologías han cambiado los medios, costos y escalas de tiempo asociadas con el manejo de datos (Tabla II). Esto se traduce en un beneficio directo para la investigación científica como la posibilidad de hacer accesible repositorios de información o bases de datos científicos a un costo mínimo, comparado con los beneficios obtenidos, como pudiera ser el incremento de la producción científica. Esto se debe a que los problemas que los

investigadores puedan resolver dependen en gran medida de la información con la que cuentan.

Tabla II. Sumario de tendencias tecnológicas que influyen en el intercambio de datos e información científica.

Tendencia	Descripción
Disminución en el costo del poder de cómputo y telecomunicaciones	El costo del poder de cómputo, almacenamiento de datos y comunicación se ha reducido en forma constante desde hace 25 años. Algunos países se han beneficiado con esto, pudiendo adquirir en forma relativamente fácil equipo moderno de cómputo y de telecomunicaciones. Los usuarios nuevos no han tenido que realizar grandes desembolsos de capital.
Mejora de las capacidades para la adquisición de datos científicos y otros datos	El poder de recolección de sus instrumentos hacen posible que grandes proyectos científicos como el Human Genome Project, el modelado del clima, estudios de percepción remota por satélite, entre otros, puedan generar grandes volúmenes de datos.
Explotación de la redes de banda ancha	La inversión en cables de fibra óptica a lo largo de las últimas dos décadas se está aprovechando cada vez más en nuevas aplicaciones que requieren de alta capacidad o entrega en tiempo real (vídeo, imágenes médicas). La industria del entretenimiento y nuevas aplicaciones como teleconferencia, vídeo sobre demanda y la televisión interactiva han atraído una gran inversión y serán los factores dominantes en el desarrollo de las redes en los próximos 10 años. La comunicación por voz requerirá una mínima porción de la capacidad de telecomunicaciones.
Advenimiento de las comunicaciones digitales inalámbricas	Los sistemas terrestres de comunicación inalámbrica han creado una moderna infraestructura en ciudades que habían contado con sistemas telefónicos poco confiables y con capacidad inadecuada. Las nuevas propuestas en comunicación satelital proveerán conexiones globales de voz y datos.
Mayor facilidad para el trabajo colaborativo	Los equipos de científicos (distantes unos de otros) pueden colaborar en proyectos de investigación gracias al alto desempeño de las telecomunicaciones en interacciones en tiempo real utilizando datos y otros recursos.
Creciente capacidad para el procesamiento del lenguaje natural	Máquinas utilizando las técnicas de procesamiento de lenguaje natural ayudan a organizar la vasta cantidad de información disponible en forma electrónica. Las nuevas herramientas proveen acceso transparente a personas que hablan idiomas diferentes.
Aumento del reconocimiento de la importancia de los estándares	Los estándares proveen los medios para la interoperabilidad, apoyan la competitividad y la evolución de productos. El reconocimiento del papel de los estándares (ya sean de facto, apoyados por la industria, por organismos nacionales o internacionales) ha crecido, acelerando la aceptación y aplicación de estándares.
Creciente aceptación de la necesidad de cooperación en el monitoreo y control de las actividades en la red.	Se han construido los mecanismos en los sistemas de autenticación, recuperación y redes para monitorear actividades específicas de los usuarios. La tecnología de codificado por llaves públicas es cada vez más aceptado para la protección como un medio para la protección de los datos y la autenticación a los usuarios.

### **III.3. Bancos de datos científicos.**

Una de las principales motivaciones de esta tesis es la de promover la creación de un repositorio de datos e información relacionada al fenómeno El Niño. La finalidad del banco de datos científicos es proveer a los investigadores del país de un repositorio central en el cual puedan tener acceso a los datos colectados y a los resultados obtenidos por los distintos grupos de investigación de este fenómeno a nivel nacional facilitando de esta manera el intercambio de información. Otra de las finalidades de la base de datos es el facilitar y promover la colaboración formando una red distribuida de investigadores reunidos en torno a ésta y proveer las herramientas de software que habiliten el acceso remoto a ella y que a la vez proporcione soporte a actividades de colaboración.

Existen diversas ventajas adicionales en el desarrollo de bases de datos científicas. La primera es que conducen al incremento de la calidad de los datos. Un número mayor de usuarios proveen múltiples oportunidades para detectar y corregir inconsistencias en los datos. Una segunda ventaja es el costo, es más económico almacenar los datos que recolectarlos de nueva cuenta. Ciertos tipos de datos como los datos ambientales son imposibles de obtener de nuevo debido a factores tan complejos e incontrolables, como el clima, la influencia de la población y los procesos dinámicos propios del ecosistema. Sin embargo, la razón principal para el desarrollo de bases de datos científicas son los nuevos tipos de investigaciones que se pueden hacer, por citar algunos: estudios a largo plazo que dependan de la base de datos para preservar la historia del proyecto, síntesis en la cual se combinen datos para otra finalidad de la que fueron originalmente recolectados y proyectos

multidisciplinarios que dependen de la base de datos para compartir la información [NRC, 1997]

Algunas bases de datos se especializan sólo en un tipo de datos específicos e implementan técnicas avanzadas para la búsqueda, recuperación y análisis. Un ejemplo es la base de datos GenBank, principal repositorio de secuencias genéticas del programa internacional Human Genome [Watson, 1990]. Esta base de datos tiene almacenados casi tres millones de secuencias de DNA [NCBI, 1998]. Otros ejemplos de bases de datos altamente especializadas son la bases de datos Protein Data Bank y a menor escala BIOTA [BNL, 1998 y UCONN, 1998].

Por otra parte, se tienen bases de datos que se utilizan en un proyecto determinado de investigación multidisciplinario. Estas incluyen una amplia gama de datos enfocados hacia un campo de estudio de la ciencia, tratando de almacenar la mayor cantidad de datos relacionados a ese campo. Un ejemplo es el National Geophysical Data Center (NGDC) que almacena datos ambientales en los campos de la geología marina y geofísica, paleoclimatología y glaciología [NOAA, 1998]. Un segundo ejemplo es el proyecto Long Term Ecological Research, cuyas bases de datos contienen datos de distinta índole como clima, movimiento de nutrientes, concentraciones de materia orgánica y biodiversidad [LTER, 1998].

En el caso del estudio del fenómeno de El Niño los requerimiento de datos que ayuden a su descripción incluye observaciones de distinta índole [IAI, 1997]:

1. Interface Atmósfera/Océano: temperatura del aire, humedad específica, niveles de presión del mar, viento, precipitación y radiación solar.
2. Altitud: vientos, temperatura del aire, presión y humedad en niveles estándares.
3. Hidrológica: flujo de ríos, niveles de lagos y reservas.
4. Oceánica: datos de perfil termal obtenidos con boyas y boyas a la deriva, estadísticas pesqueras y biología marina.
5. Recursos marinos renovables: registro de plancton y series de datos de clorofila en estaciones fijas y cruceros, series selectas de especies indicadoras, inventarios de especies existentes y su condición.

#### **III.4. Técnicas de almacenamiento de datos científicos.**

El manejo de datos científicos presenta algunos retos. Entre los principales están el almacenamiento y recuperación. Los datos científicos se caracterizan principalmente por la rica variedad de tipos de datos que utilizan (i.e. imágenes, arreglos multidimensionales) y estructuras que se requieren, el tamaño de esas estructuras y la información adicional que se debe de incluir con ellas. A esto hay que agregar los requerimientos necesarios para procesar los datos almacenados, por ejemplo un buen desempeño en funciones de entrada/salida ó un manejo eficiente del espacio de almacenamiento que facilite intercambio, entre otros.

En aplicaciones comerciales es común la utilización de sistemas manipuladores de bases de datos, principalmente sistemas de bases de datos relacionales [Ullman, 1988]. Estos resultan poco prácticos en el manejo de datos científicos, ya que generalmente los datos que se almacenan son datos multidimensionales y/o jerárquicos y presentan formatos muy

variados. Los sistemas de bases de datos relacionales están diseñadas para proveer acceso detallado a relativamente pocas estructuras de datos bien definidas y no soportan objetos multidimensionales. Adicionalmente estos sistemas son poco eficientes para el acceso secuencial a volúmenes de datos requeridos. En un futuro, posiblemente las bases de datos orientadas a objetos puedan dar solución a estos problemas [Kim, 1990].

Para solucionar el problema del almacenamiento de datos científicos, se han desarrollado distintos formatos de archivos orientados en forma especializada a estos tipos de datos. Existen diferentes formatos “estándar” para el manejo de datos científicos. Los formatos más utilizados en series de datos para datos atmosféricos y oceanográficos son:

1. GRIB (GRId in Binary). Es un formato binario empaquetado para transmitir y archivar grandes volúmenes de datos oceanográficos y meteorológicos en arreglos bidimensionales. Es el estándar utilizado por dos de los principales centros meteorológicos a nivel mundial, el U.S. Climate Prediction Center y el European Centre for Medium-Range Weather Forecast [WOM, 1994].
2. CDF (Common Data Format). Desarrollado por la NASA en 1980 a manera de interface y utilerías de software que permiten el almacenamiento y acceso a datos en arreglos multidimensionales, almacenados en máquinas VAX a través de programas FORTRAN. Actualmente es un formato independiente de la plataforma y es utilizado por diferentes grupos de la NASA [NSSDC, 1998] .

3. netCDF (network CDF). Es una interface para el acceso a datos científicos. Fue desarrollado por UCAR (University Corporation for Atmospheric Research) alrededor de 1986 utilizando CDF como base. Las librerías de CDF se modificaron para proveer recuperación y almacenamiento en forma independiente de la plataforma. netCDF es incompatible con el formato CDF y en la actualidad el software para su conversión está en etapa de desarrollo. NetCDF utiliza una interface común para el acceso a los datos implementada sobre una representación de los datos que es independiente de la arquitectura. El acceso a los datos en los archivos puede ser secuencial o aleatorio [UCAR, 1998].
4. HDF (Hierarchical Data Format). Es un formato genérico y extensible para el intercambio de datos científicos desarrollado por el National Center for Supercomputing Applications (NCSA). Se basa en un formato común para los datos sobre el cual se construyen distintas interfaces de aplicación para el acceso a los datos [NCSA, 1998].
5. BUFR (Binary Universal Format Representation). Es un formato para la representación de datos meteorológicos y oceanográficos. Aunque pueden ser usados para almacenar cualquier otro tipo de datos, su principal función es representar datos observacionales. Fue diseñado para eliminar la redundancia y hacer más eficiente su transmisión en forma electrónica y reducir el tiempo de procesamiento requerido para decodificar la información [WOM, 1994].

La razón de que existan tantos estándares es en parte histórica y práctica. Muchas organizaciones desarrollaron sus propios formatos para el almacenamiento de datos, cuando

otros grupos les solicitaban información, se enviaban los datos en su propio formato, lo que ocasionó que distintos grupos utilizaran un formato en particular de forma común. Así es como estos formatos se volvieron *estándares de facto*. Por otra parte, el desarrollo de cualquier formato de almacenamiento e intercambio implica inclinarse por alguna característica o requerimiento como el tamaño de los archivos, simplicidad, facilidad de intercambio y transportabilidad. De esta manera, algunos formatos son más eficientes para el almacenamiento y transmisión, mientras que otros soportan estructuras más accesibles. Por ejemplo, netCDF requiere que el tamaño de sus campos de información sean múltiplos de 8 bits, a diferencia del formato GRIB que no tiene esta restricción. Un tipo de dato que pudiera ser representado de manera más eficiente con nueve bits requeriría casi el doble de espacio de almacenamiento en netCDF. Sin embargo, una de las ventajas de netCDF es que es un formato autodescribible y es posible agregar una gran variedad de metadatos. A continuación se presenta una comparación de los tres formatos principales para almacenamiento de datos científicos (Tabla III).

Tabla III. Resumen de características de los formatos para datos científicos CDF, HDF y netCDF.

	CDF	HDF	netCDF
Lenguaje	C, Fortran	C, Fortran,Java	C, Fortran,C++
Tipos de datos	char, short, int, float, double, string	byte, short, long, float, double string	byte, char, short, long, float, double
Tipos de datos definidos por el usuario	✗	✓	✗
Método de conversión	XDR, Nativo	XDR, Nativo	XDR
Máxima dimensión de arreglos	10	Ilimitada	32
Dimensión extensible de arreglos	✓	✓	✓
Atributos definidos por el usuario	✓	✓	✓
Tipos de atributos	Cualquiera	Cualquiera	Cualquiera
Identificadores de dimensiones	✗	✓	✓
Orden de índices en arreglos	Renglón, Columna	Renglón, Columna	Renglón, Columna
Compartido	✓	✓	✓
Compresión	✗	✓	✗
Herramientas	pocas	muchas	limitadas

### III.5. El formato de archivo HDF.

El formato que se propone utilizar para este trabajo es el HDF [NCSA, 1993]. HDF es un formato desarrollado para facilitar a los investigadores y programadores el almacenamiento, transferencia y distribución de datos creados en varias máquinas con diferente software. Adicionalmente, el formato HDF es también una colección de software, interfaces de aplicación y utilerías que constituyen las librerías de HDF y permite a los usuarios trabajar con archivos HDF.

El formato HDF es un formato de archivo multiobjeto para compartir información científica en un ambiente distribuido. Fue diseñado para satisfacer las necesidades de grupos

de investigadores trabajando en proyectos de varios campos. Para su diseño se consideraron lo siguientes requerimientos [NCSA, 1997]:

1. Proveer mecanismos para que los programas obtengan la información acerca de los datos en un archivo dentro del mismo archivo.
2. Permitir al usuario almacenar combinaciones de datos de distintas fuentes en un mismo archivo, así como almacenar los datos y su información relacionada en archivos separados aun cuando los archivos son procesados por la misma aplicación.
3. Estandarizar el formato y descripciones de conjuntos de datos comúnmente usados como imágenes raster y arreglos multidimensionales.
4. Puede ser adaptado para dar cabida a cualquier tipo de dato.
5. Ser independiente de la plataforma.

Los archivos HDF son autodescribibles, lo que significa que para cada estructura de datos en un archivo HDF se incluye una amplia información acerca de su estructura y su localización en el archivo. A esta información se le conoce como metadatos. Adicionalmente, se pueden incluir múltiples tipos de datos en un mismo archivo HDF utilizando las estructuras apropiadas. Estas estructuras reciben el nombre de *objetos de datos* y están formados por un elemento descripción (descriptores) y de los datos. Los descriptores de datos (Figura 8) contienen información acerca del tipo, localización y tamaño de un elemento de datos. Los elementos de datos contienen los datos mismos.

Etiqueta	Referencia	Desplazamiento	Longitud
2 bytes	2 bytes	4 bytes	4 bytes

Figura 8. Contenido del descriptor de datos en archivos HDF.

Las etiquetas identifican el tipo de dato almacenado en su elemento de datos. Por ejemplo una imagen raster tendrá en éste campo la etiqueta DFTAG\_RI. Los números de referencia ayudan a distinguir aquellos elementos de datos que tengan la misma etiqueta. La unión de la etiqueta y el número de referencia forman un identificador único para cualquier objeto almacenado en un mismo archivo HDF. El campo de desplazamiento indica la localización del elemento de datos en el archivo almacenado, el número de bytes que hay desde el inicio del archivo hasta el inicio del elemento de datos. La longitud indica el tamaño total del elemento de datos en bytes.

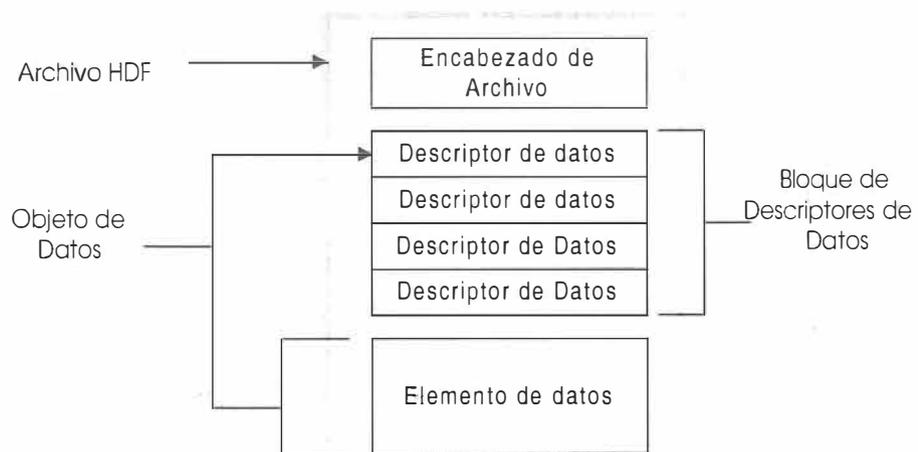


Figura 9. Organización física de un archivo HDF

HDF puede ser descrito a distintos niveles de abstracción. En su nivel más bajo, HDF es un formato de archivo físico para almacenar datos científicos. A su nivel más alto, HDF es una colección de utilerías y aplicaciones para manipular, visualizar y analizar los datos almacenados en un archivo HDF. Entre estos dos niveles, se encuentra una librería formada

por dos capas o interfaces: la interfaz de programación de bajo nivel y la interfaz de programación de alto nivel o de aplicación

En lo referente a las funciones de bajo nivel, ésta interfaz incluye el manejo de flujos de entrada/salida, manejo de errores, administración de memoria y almacenamiento físico. Esta capa es para el uso de usuarios con experiencia que deseen extender la funcionalidad de la interfaz de alto nivel.

La interfaz de alto nivel está formada por varios conjuntos de rutinas independientes diseñadas para simplificar el proceso de almacenamiento y acceso de un tipo de dato en particular. La mayoría de estas rutinas están disponibles en FORTRAN-77 y C.

Otro factor crítico en el manejo de bases de datos científicas es el acceso o recuperación de los datos en forma distribuida. Los datos comúnmente están almacenados en archivos grandes y con una estructura muy compleja. Al acceder estos archivos, los usuarios por lo regular sólo requieren de una porción específica del conjunto total de datos. Es por eso que es de vital importancia proporcionar a los usuarios los medios necesarios para identificar los datos requeridos y recuperar sólo la porción de datos seleccionada en un formato utilizable.

### **III.6. Tecnologías de acceso remoto a bancos de datos científicos.**

Actualmente algunas bases de datos científicas existentes tienen implementados sistemas servidores de datos científicos, orientados a proporcionar acceso a los datos almacenados a través de la World Wide Web (WWW). El problema que presenta este

enfoque es en primer lugar la poca interactividad que proporcionan el modelo consulta/respuesta del protocolo HyperText Transfer Protocol (HTTP) utilizado en WWW [Tannebaum, 1996]. Otro problema es que los navegadores de paginas WWW (browsers) no soportan el despliegue de la mayoría de los archivos para datos científicos. Existen algunas implementaciones que incluyen tecnologías como CGI o Servlets en la parte del servidor para aumentar la interacción entre el servidor y los clientes, lo cual funciona bien pero con un desempeño muy pobre. Lo anterior se debe principalmente a la latencia de la red y sobrecarga del servidor.

Cuando se interactúa con datos científicos, las operaciones más usuales son examinar el contenido de los archivos, seleccionar algún subconjunto específico del total de datos almacenados, transformar o hacer combinaciones con esos datos o visualizar algún objeto específico. Es por esto que es necesario que el servidor no se limite sólo al envío de los datos.

La solución que se propone en este trabajo para resolver este problema, consiste en el diseño e implementación de un sistema cliente/servidor para la visualización de datos científicos en forma distribuida que cumpla con los siguientes requerimientos:

- Permita la identificación de los atributos y estructura de los datos almacenados en un archivo con formato HDF en particular, sin tener que recurrir a la transferencia del archivo completo.

- Permita la selección de objetos particulares almacenados y facilite su transferencia a través de la red. Lo que se traduce en un uso eficiente de los recursos disponibles (i.e. conexiones de red y memoria principal).
- El acceso a los archivos se realizará en forma transparente. En lo posible el acceso a los archivos almacenados en la bases de datos remota, deberá de realizarse en forma similar al acceso a archivos en el sistema de archivos local del usuario. El sistema deberá permitir el acceso a archivos almacenados en distintos servidores de datos.
- Deberá permitir compartir los datos almacenados en un mismo archivo en sesiones de trabajos colaborativas. Debe también permitir el acceso a múltiples archivos de datos en una misma sesión de trabajo.

Para la implementación de este sistema, se seleccionó el lenguaje de programación Java [Gosling, 1996]. Este lenguaje fue diseñado con la finalidad de solucionar muchos de los problemas que surgen en el tipo de aplicaciones que se desarrollan actualmente. Inició como parte de un proyecto mayor que tenía por objetivo el desarrollo de una ambiente para el desarrollo de software instalado en aparatos electrónicos. Java es un lenguaje interpretado para la programación orientada a objetos. Las principales características por las que lo se selecciono este lenguaje son las siguientes:

1. Es un lenguaje independiente de la arquitectura. Desde sus inicios Java fue diseñado para el desarrollo de aplicaciones para ambientes de redes de computadoras. En general, estas redes incluyen una gran variedad de arquitecturas tanto de hardware

cómo de software. Para permitir que las aplicaciones de Java puedan correr en cualquier máquina, los compiladores generan un archivo de código “objeto” independiente de la plataforma, el cual es ejecutable en cualquier computadora que contenga el ambiente de tiempo de ejecución o Máquina Virtual de Java. Esta es una gran ventaja, ya que permitirá que el sistema implementado pueda ser utilizado en cualquier computadora que esté al alcance de los usuarios.

2. Facilidad para desarrollo de aplicaciones distribuidas. Java cuenta con una gran cantidad de librerías de clases para el manejo de los protocolos TCP/IP (i.e. UDP, TCP, HTTP) ampliamente utilizados en Internet y de abstracciones de alto nivel (i.e. sockets, streams, threads) para la creación de sistemas bajo el modelo cliente/servidor.

Una parte del presente trabajo consiste en el desarrollo de un mecanismo para el acceso remoto a archivos almacenados en una máquina remota. Para lo cual se explora la aplicación de una tecnología incorporada recientemente al lenguaje Java para el desarrollo de sistemas basados en objetos distribuidos. Esta tecnología se denomina Remote Method Invocation (RMI) y permite a los programas de Java acceder a métodos (procedimientos o funciones) de objetos remotos [Sun, 1998]. En el siguiente capítulo se presenta una descripción más detallada de esta tecnología.

## CAPÍTULO IV DISEÑO E IMPLEMENTACIÓN

Se presenta la descripción de los diferentes prototipos que se construyeron en el proceso de desarrollo del sistema. En total se implementaron tres prototipos, según el modelo de desarrollo en espiral propuesto por Bohem [1988] y en el paradigma de desarrollo de software orientado a objetos [Booch, 1994]. Esta combinación permite el refinamiento sucesivo de la arquitectura del sistema en forma iterativa y se añade mayor funcionalidad a cada uno de los prototipos que se construyen a lo largo del todo el proceso de desarrollo. Para el modelado del sistema se utiliza el Lenguaje Unificado de Modelado (Unified Modeling Language, UML). Como se mencionó anteriormente, la caracterización del espacio de diseño del sistema colaborativo se realizó conforme a la metodología propuesta por Zapata [1997].

La plataforma de herramientas de desarrollo utilizada en este trabajo incluye:

- Java Development Kit (JDK) 1.1.6 y 1.1.7
- Java Foundation Classes (a.k.a. Swing) 1.0.3 y 1.1beta3
- NCSA HDF Library 4.1r2
- NCSA JHI 2.0

En cuanto a la plataforma de hardware se utiliza una computadora con procesador 486 de 100Mhz, 32Mb de RAM y disco duro de 500Mb corriendo Windows 95.

## **IV.1. Desarrollo del primer prototipo.**

### **IV.1.1. Java y las librerías de HDF.**

La funcionalidad del primer prototipo se centró en la creación de una sistema visualizador de archivos con formato HDF que permita el acceso a archivos almacenados en el sistema de archivos local. La construcción de este prototipo permitió en primer lugar la familiarización con las librerías para el manejo de archivos HDF desarrolladas por NCSA, en este caso la versión utilizada de estas librerías fue la 4.1r2. Debido a que las librerías fueron desarrolladas en Lenguaje C, es necesario utilizar una interface que permita la comunicación de estas librerías con código en Java. Debido a la creciente popularidad de éste lenguaje, NCSA creó Java HDF Interface (JHI) que es un paquete de clases (wrapper classes) que encapsulan las librerías en código nativo. Es importante hacer énfasis en que ésta no es una implementación en Java de la librería de HDF, sino que se hace llamado a las mismas librerías que se utilizan en los programas codificados de C o FORTRAN. La interface JHI está compuesta por paquetes de clases en Java (i.e. ncsa.hdf.hdfplib) y librerías de código nativo (i.e. hdf.dll y jhv.dll para la versión Windows 95/NT) enlazadas en forma dinámica. En las clases se declaran métodos nativos y las librerías contienen las funciones en C que implementan estos métodos nativos.

En el paquete de distribución de JHI se incluye un visualizador de archivos HDF llamado Java HDF Viewer (JHV). Está implementado como un aplicación de Java, construida sobre JHI y las librerías de HDF. Adicionalmente se implementan clases para el

despliegue jerárquico de objetos almacenados en archivos HDF, despliegue de metadatos, anotaciones, datos e imágenes (Ver Figura 10).

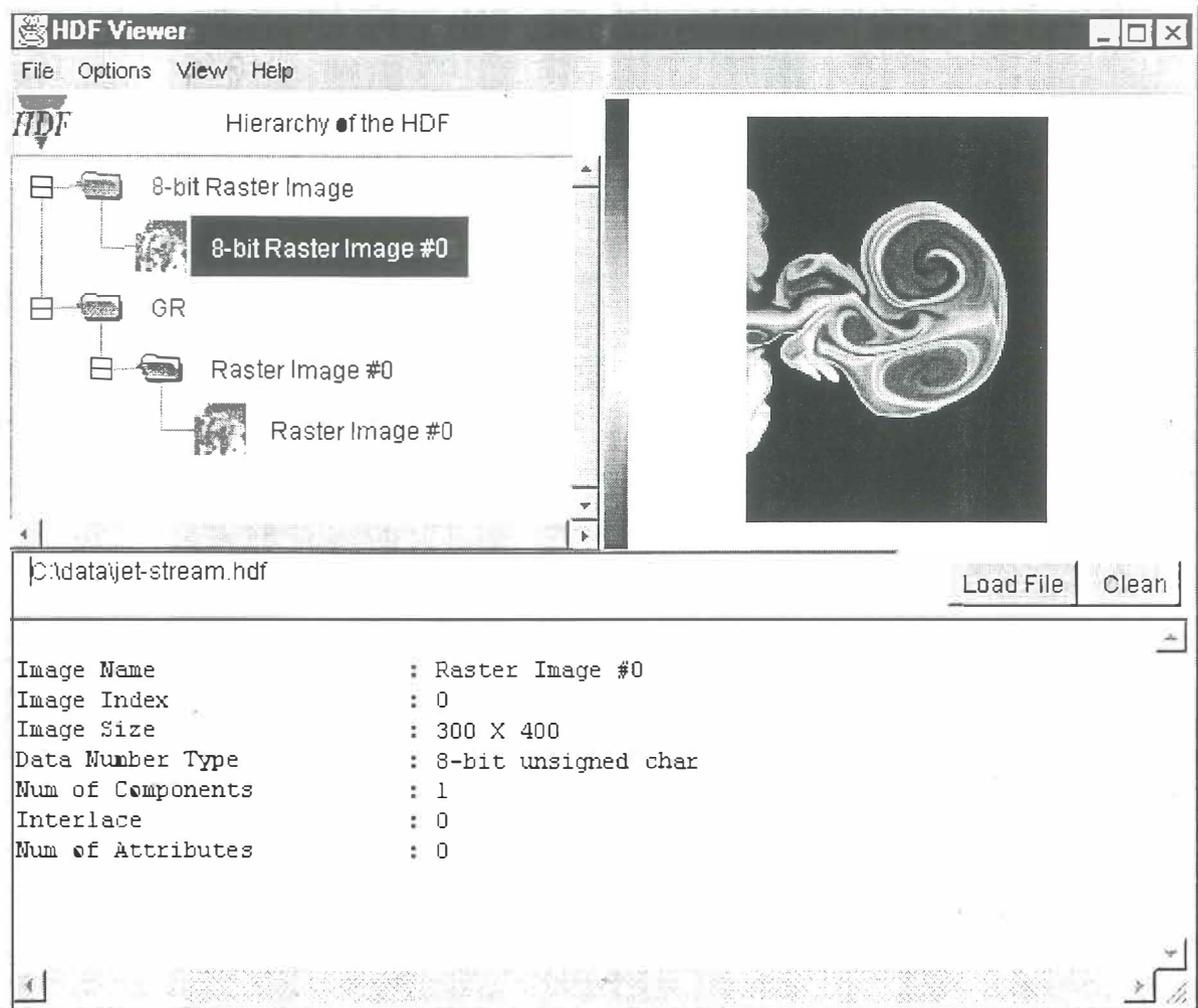


Figura 10. Ventana principal de la aplicación Java HDF Viewer (JHV).

#### IV.1.2. Descripción de la arquitectura del prototipo.

Este prototipo se construyó aprovechando la interface JHI para el acceso a las librerías de HDF. La Figura 11 muestra arquitectura general del prototipo.

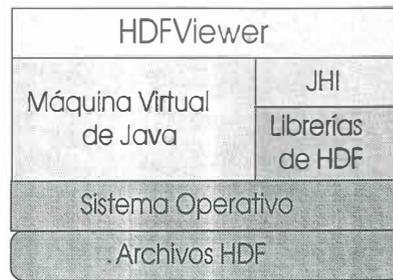


Figura 11. Arquitectura del primer prototipo

La capa ubicada en la parte superior de la arquitectura (HDFViewer) es la aplicación que se desarrolló en este trabajo, la cual comprende los distintos componentes de la interface de usuario necesarios para la manipulación y representación de los objetos almacenados en los archivos, así como los mecanismos para la recuperación de estos objetos. Para lo anterior, los componentes de esta capa utilizan los servicios de la interface JHI para el acceso a las librerías de HDF. La Figura 12 muestra el diagrama de objetos que describe las relaciones que guardan las clases que constituyen este prototipo.

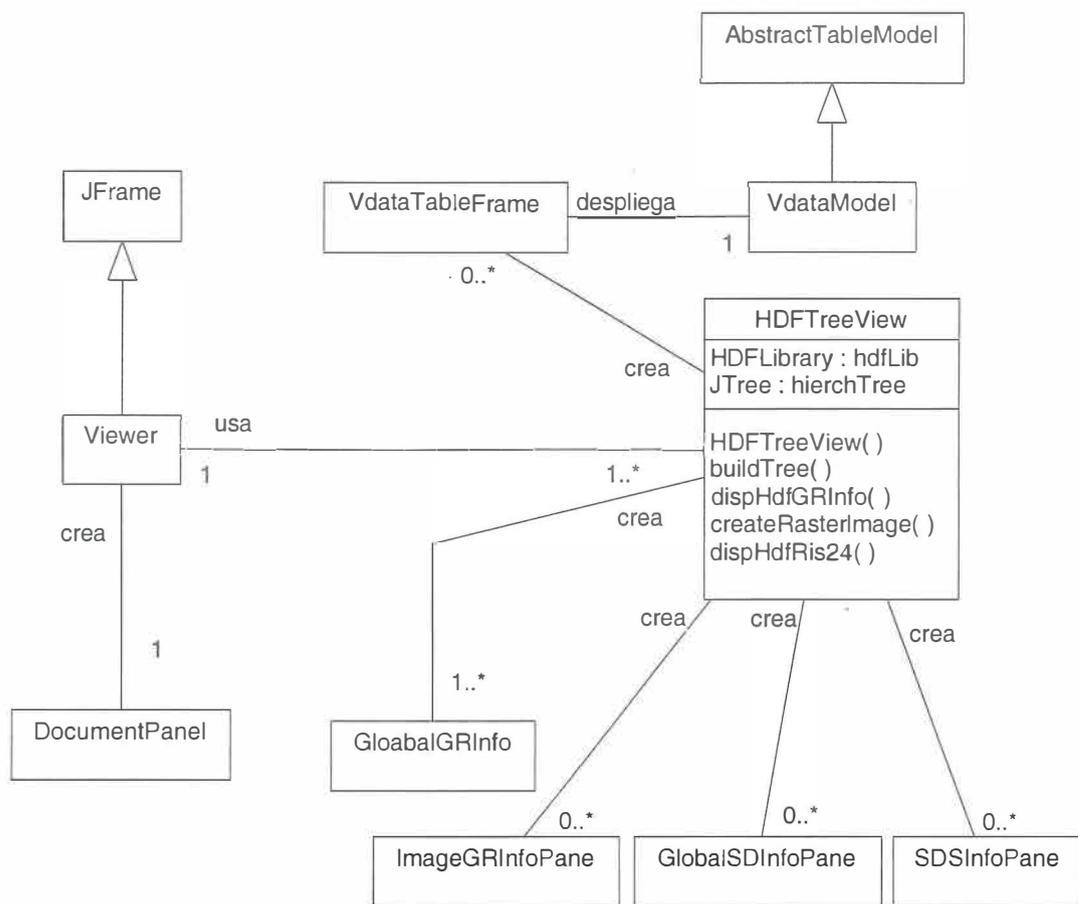


Figura 12. Modelo de objetos del primer prototipo.

El modelo de objetos resultó simple. Esto se debe en gran parte a que se aprovecharon muchos de los componentes para la creación de interfaces gráficas incluidos en el Java Foundation Classes (JFC), de esta manera el diseño estuvo más enfocado en la parte del mecanismo de recuperación de los objetos. La Figura 13 presenta el diagrama de secuencia que describe éste mecanismo.

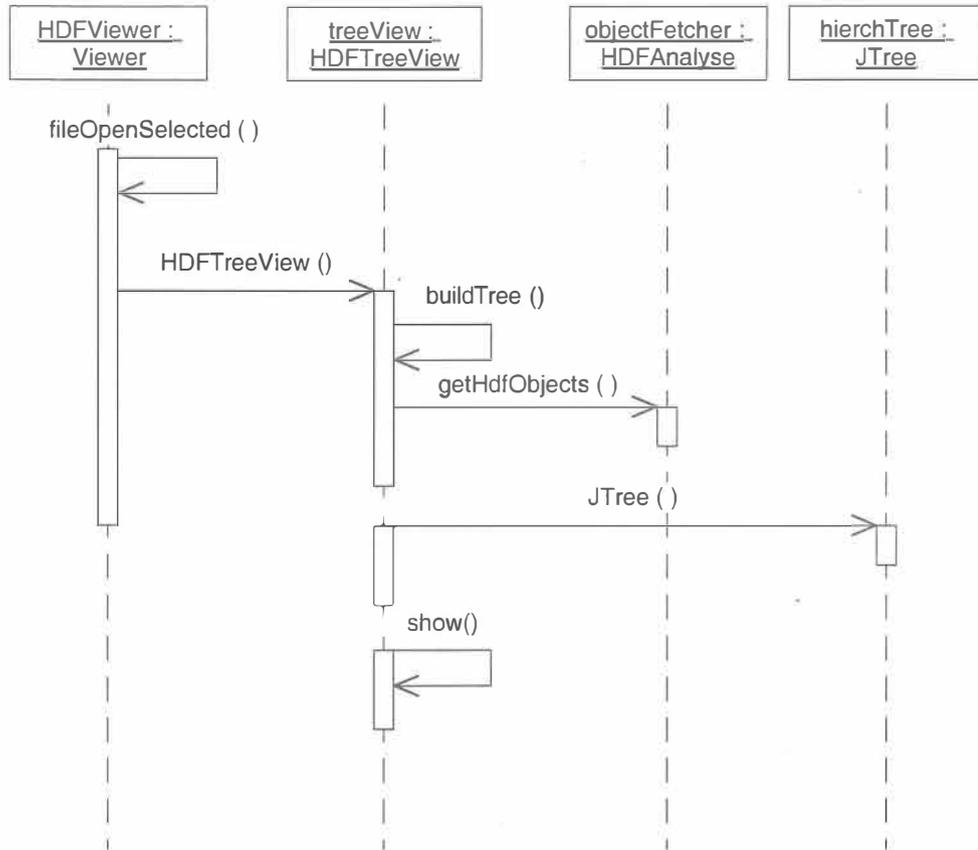


Figura 13. Diagrama de secuencia para la recuperación de objetos de datos en un archivo HDF.

Como se mencionó en el capítulo anterior, todo objeto almacenado en un archivo HDF tiene asociado un descriptor de datos que permite su identificación dentro de ese archivo. Como primer paso, es necesario recuperar estos descriptors para posteriormente utilizarlos para recuperar sus elementos de datos asociados. Este proceso se inicia cuando el usuario selecciona la opción de abrir un archivo en el menú principal de la aplicación (HDFViewer en la Figura 13) lo que ocasiona la invocación del método `fileOpenSelected` de ese mismo objeto. Una vez que se ha seleccionado el archivo deseado, se crea un nuevo objeto de la clase `HDFTreeView`, la cual tiene como responsabilidades la representación visual de la jerarquía de objetos almacenados en un archivo HDF y almacenar los descriptors de datos asociados a esos objetos para su futura

recuperación. Cuando se crea un nuevo objeto de la clase `HDFTreeView`, se hace el llamado al método `buildTree`, que es el encargado de construir la jerarquía de objetos auxiliándose de la clase `HDFAnalyse` definida en JHI. Adicionalmente, JHI proporciona la clase `HDFObjectNode` la cual encapsula la representación de los descriptores de datos. El método `getHdfObject` de la clase `HDFAnalyse` obtiene cada una de los descriptores de datos contenidos en un archivo y los almacena en una cola de objetos `HDFObjectNode`. A partir de esta cola `buildTree`, construye un árbol con la jerarquía de objetos del archivo HDF. La ventaja de éste enfoque consiste en que sólo se almacenan las referencias a los objetos contenidos en el archivo. De esta forma, la clase `HDFTreeView` nos sirve al mismo tiempo de componente visual que provee la información acerca del contenido de un archivo, y también como índice para la recuperación de los datos. En la Figura 14 se puede apreciar el resultado final de este proceso.

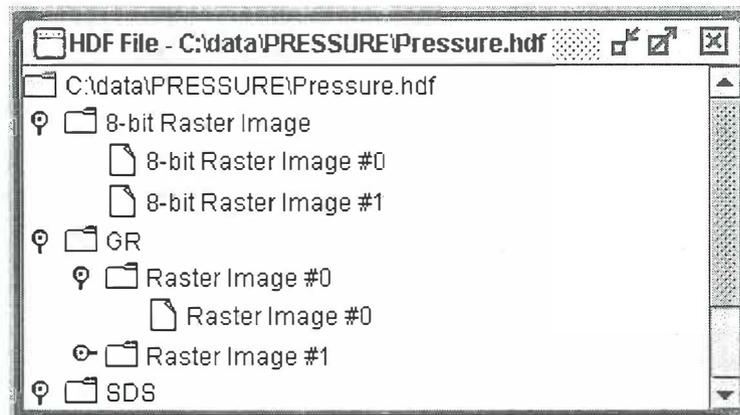


Figura 14. Componente `HDFTreeView`, representación gráfica de la jerarquía de objetos de un archivo HDF.

### IV.1.3. Interface de usuario.

La interface de usuario del sistema se basa en el modelo de Interface de Documentos Múltiples (MDI, por sus siglas en inglés). Este tipo de interface gráfica permite acomodar

varias subventanas dentro de una ventana principal, la cual se convierte en un nuevo *desktop* sobre el que operan las subventanas. Por ejemplo, cuando se minimiza una subventanas esta se convierte en un icono que aparece en la parte inferior de la ventana principal. Siguiendo este modelo, el sistema permite tener varios archivos HDF abiertos a un mismo tiempo. A cada uno de estos archivos se asocia un componente `HDFTreeView` (ver Figura 14) mediante el cual el usuario tiene acceso a los datos almacenados en el archivo. Cada uno de los nodos del árbol tiene asociado un objeto en su respectivo archivo. El contenido del archivo `pressure.hdf` incluye, entre otros datos, dos imágenes raster de 8 bits, nodos etiquetados como `8-bit Raster Image #0` y `8-bit Raster Image #1`. Para recuperar alguna de estas imágenes, el usuario sólo tiene que seleccionar la imagen deseada dando un “doble click” sobre el nodo que identifica a la imagen y ésta se desplegará en otra ventana. La Figura 15 presenta la ventana principal del sistema. En este caso se tienen 5 archivos abiertos representados por su respectivo árbol jerárquico, tres de los cuales se encuentran minimizados. Adicionalmente se tiene el despliegue de una imagen raster, y se puede ver como en la parte inferior de la ventana principal se agrupan las ventanas minimizadas. La barra de menú cuenta con un sólo elemento, esto se debe a que la mayor parte de la interacción y el sistema se hace a través de la manipulación de los árboles jerárquicos asociados a cada archivo.

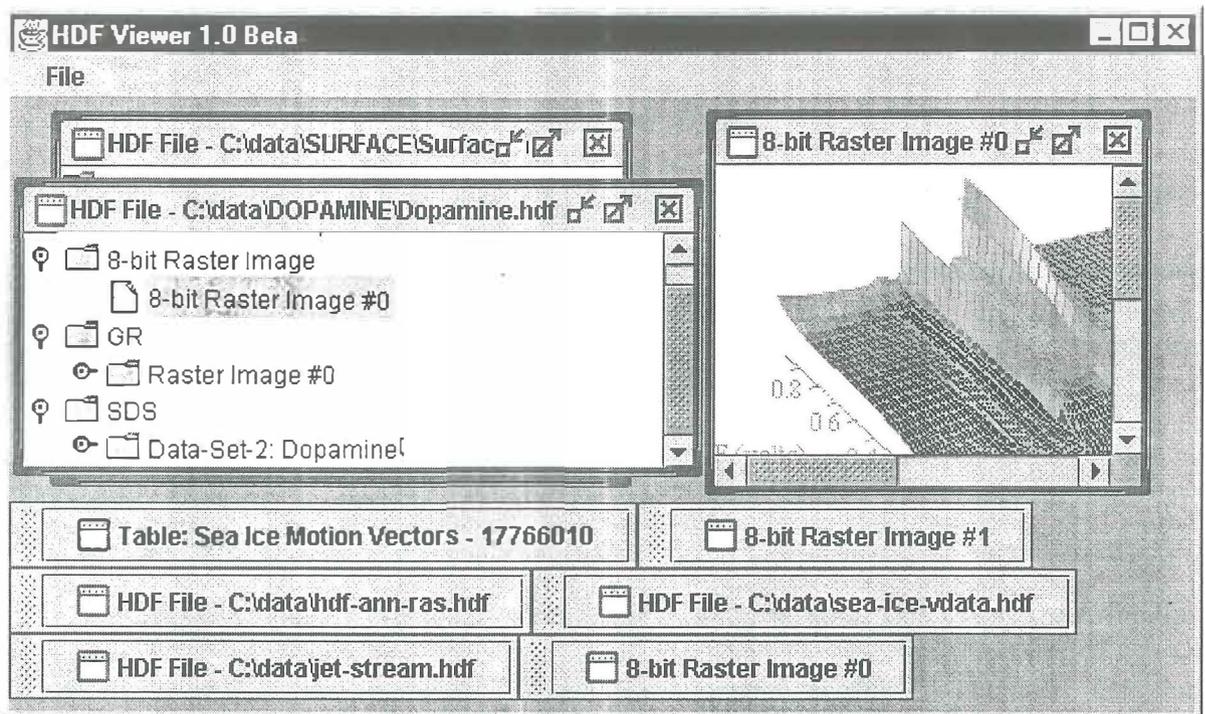


Figura 15. Interface de usuario del primer prototipo.

## IV.2. Desarrollo del segundo prototipo.

En esta sección se presenta el desarrollo del segundo sistema prototipo. Para su elaboración se retomó el prototipo anterior, añadiéndole funcionalidad para el acceso distribuido a los archivos HDF. El sistema se basa en la arquitectura *cliente/servidor*. Esta arquitectura comprende dos partes: 1) un servidor que provee algún servicio y 2) cliente que solicita los servicios al servidor. Los clientes implementan la funcionalidad requerida por los usuarios y los servidores se encargan de prestar los servicios necesarios para proveer esa funcionalidad [Lewandowski, 1998].

La implementación de esta arquitectura se realizó mediante la utilización de los mecanismos para la creación de *objetos distribuidos* del lenguaje Java. El paradigma de orientación a objetos ha ganado terreno al enfoque de procedimientos en el desarrollo de

software. La programación orientada a objetos es un método de implementación en el cual los programas están organizados como colecciones cooperativas de objetos, cada uno de los cuales representa una instancia de una clase, y las clases son parte de una jerarquía de clases unidas por medio de relaciones de herencia [Booch, 1994]. Un objeto puede ser visto como una entidad que encapsula ciertos datos (el estado del objeto) y ofrece un conjunto de operaciones que actúan sobre esos datos (comportamiento del objeto). Comúnmente estas operaciones reciben el nombre de *métodos*.

Tradicionalmente los objetos residen en un solo programa. Los objetos distribuidos aumentan su espacio de operación, ya que estos pueden residir en cualquier máquina conectada a una red proveyendo de sus servicios en forma remota a otros objetos [Orfali, 1998].

#### **IV.2.1. RMI.**

Remote Method Invocation (RMI, por sus siglas en inglés) es la tecnología del lenguaje Java que permite la creación de objetos distribuidos. Fundamentalmente RMI es un mecanismo de Remote Procedure Calls (RPC por sus siglas en inglés) que incluye ventajas adicionales. Debido a que RMI es parte del modelo de objetos de Java, este puede utilizar objetos como argumentos y valores de retorno, no únicamente tipos de datos predefinidos. Esto significa que es posible enviar como un solo argumento tipos complejos (i.e. un objeto de la clase Vector de Java). En los sistemas de RPC existentes se tendría que descomponer este objeto en tipos de datos primitivos (i.e. enteros, números de punto flotante, cadenas), enviar esos datos y reconstruir el objeto en el servidor. En el modelo de objetos distribuidos

de Java, un objeto remoto es un objeto cuyos métodos pueden invocarse desde otra máquina virtual de Java, normalmente ubicada en otra computadora. Cada objeto remoto implementa una o más interfaces remotas que especifican que métodos se pueden invocar. Estos métodos se invocan de manera muy similar a como lo harían los objetos locales. A manera de ejemplo muy general, se puede invocar un método remoto de un objeto servidor para que calcule la suma de una serie de registros en una base de datos y regrese el resultado. Lo cual es más eficiente que traer los registros y realizar la suma de manera local. El sistema de RMI esta formado por 3 capas cómo se muestra en la Figura 16.

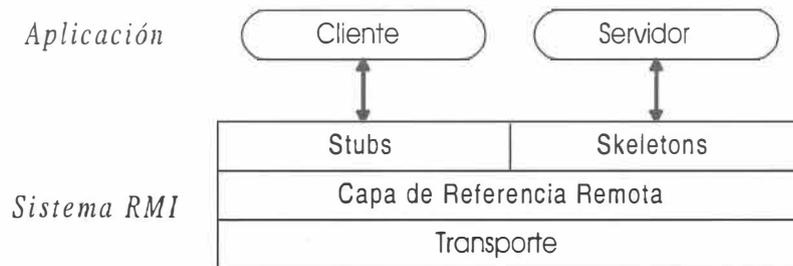


Figura 16. Arquitectura en capas del sistema RMI

El sistema de RMI se encuentra sobre la Máquina Virtual de Java y aprovecha los mecanismos de recolección de basura, seguridad y carga de clases. La invocación de un método remoto del cliente al objeto servidor recorre hacia abajo las capas del sistema RMI, hacia la capa de transporte del cliente. En esta capa se envía la invocación a la capa de transporte del servidor. La invocación se envía a la capa de referencia remota del servidor y realiza cualquier operación que sea necesaria antes de pasar la invocación al “skeleton”. El skeleton de un objeto remoto hace el llamado a la implementación del objeto remoto, la cual se encarga de ejecutar el procedimiento definido para ese método. Los valores de retorno se envían nuevamente al skeleton, el cual los envía a la capa de referencia remota y transporte

del lado del servidor, y después hacia arriba de las capas de transporte, referencia remota y stub de la máquina virtual de Java del cliente que invocó el método.

Cuando un cliente invoca un método del objeto servidor lo hace a través de un “proxy” o “stub” (objeto delegado) del objeto servidor. La referencia que tiene el cliente del objeto remoto es una referencia a un stub local, que es una implementación de las interfaces remotas del objeto servidor que remite las invocaciones al objeto servidor a través de la capa de referencia remota. La capa de referencia remota es responsable de manejar las invocaciones. Se encarga, entre otras cosas, de determinar si el servidor es un objeto único o es un objeto replicado que requiere de la comunicación con varias máquinas. La capa de transporte es responsable de la configuración y administración de la conexión, el monitoreo de conexiones activas y de la localización de los objetos remotos. Un skeleton para un objeto remoto es la contraparte del servidor, su función es despachar las invocaciones a la implementación del objeto remoto.

#### **IV.2.2. Descripción de la arquitectura del segundo prototipo.**

La Figura 17 muestra la arquitectura del sistema, el cual está compuesto por dos subsistemas, el servidor y el cliente. El servidor es capaz de atender a varios clientes en forma simultánea. Si se compara este prototipo con el primer prototipo desarrollado, se observa que ahora la funcionalidad del sistema se dividió entre el servidor y el cliente. El cliente nos proporciona la interface con la que el usuario interactúa con el sistema y se encarga de solicitar al servidor los datos que el usuario requiere. El servidor tiene la responsabilidad de atender las peticiones de los clientes, buscar los datos solicitados y

enviarlos al cliente. Ya que el servidor está encargado del manejo de los archivos, sólo en esta parte es necesario tener las librerías de HDF.

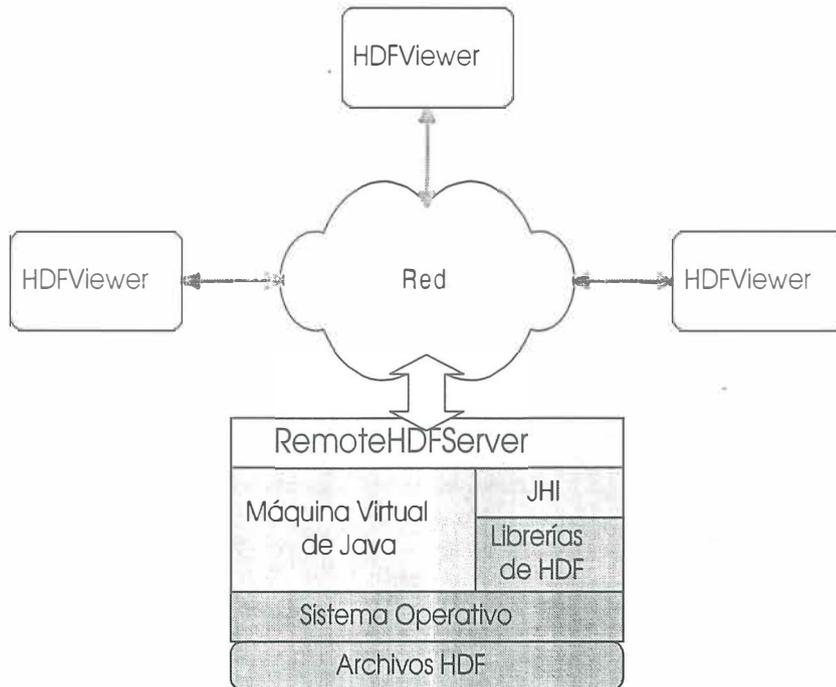


Figura 17. Arquitectura del segundo prototipo.

El servidor (`RemoteHDFServer`) está formado por dos clases. La clase `RMIFileSystemImpl` que implementa una interface remota para permitir el acceso al sistema de archivos de la máquina en donde se ejecuta el servidor. La clase `RemoteHDF` tiene la responsabilidad de proporcionar el acceso remoto al contenido de un archivo HDF específico. El diagrama de objetos del servidor se presenta en la Figura 18.

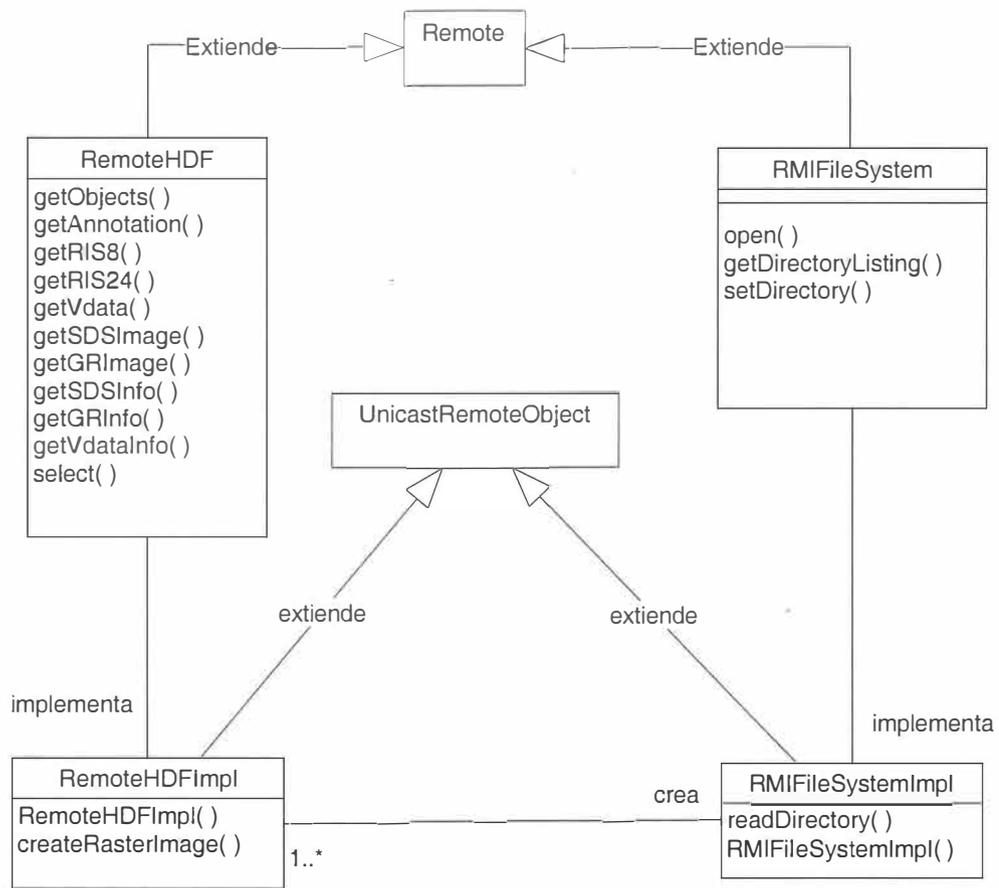


Figura 18. Modelo de objetos del sistema RemoteHDFServer.

Las clases `RMIFileSystemImpl` y `RemoteHDFImpl` extienden a la clase `UnicastRemoteObject` la cual es parte de la API de Java, e implementa un objeto servidor remoto con las siguientes características: todas las referencias a este objeto remoto son válidas únicamente durante el periodo de vida del proceso que creó el objeto remoto y requiere como mecanismo de un transporte basado en conexiones como el TCP. `RMIFileSystem` y `RemoteHDF` son las interfaces que declaran los métodos remotos implementados en `RMIFileSystemImpl` y `RemoteHDFImpl`.

La Figura 19 muestra el mecanismo de creación del servidor de datos y la activación de los servicios de acceso al sistema de archivos de la máquina en donde residen los archivos

HDF. El proceso se inicia con la instalación de un administrador de seguridad en la Máquina Virtual de Java. Esto se hace mediante la creación de una nueva instancia de la clase `RMIFileManager`, el cual se activa mediante la invocación del método `setSecurityManager` de la clases `System`. Este manejador se encarga de controlar las acciones de las clases cargadas en forma remota (i.e. stubs). El siguiente paso es crear una nueva instancia de `RMIFileSystemImpl` la cual implementa la interface remota para el acceso al sistema de archivos de la máquina. Por seguridad, sólo se tiene acceso a los directorios que se especifican en una lista predefinida en un archivo de configuración. Esta lista se almacena en memoria en una tabla y se consulta cada vez que se requiere cambiar el directorio de donde se accesan los archivos.

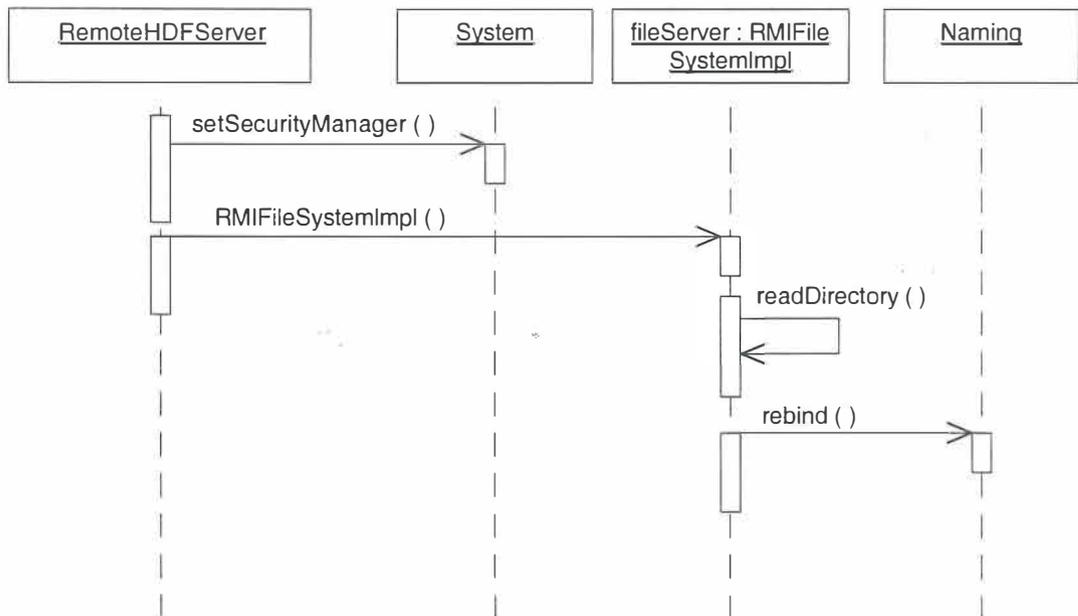


Figura 19. Diagrama de secuencia de la creación del sistema `RemoteHDFSServer`.

A continuación se lee el directorio en donde se encuentran los archivos que se pondrán disponibles para ser accedados remotamente. El método `readDirectory` guarda los

nombres de los archivos almacenados en ese directorio en un vector que se mantiene en la memoria principal. Lo anterior se hace por razones de eficiencia, ya que con esto se evita tener que acceder al dispositivo de almacenamiento cada vez que se desee abrir un archivo y en lugar de ello se accesa a la tabla que se mantiene en memoria. El último paso es dar de alta en el *registro de RMI* el nuevo objeto remoto ejecutando el método `rebind` de la clase `Naming`. La Figura 20 describe éste proceso más detalladamente.

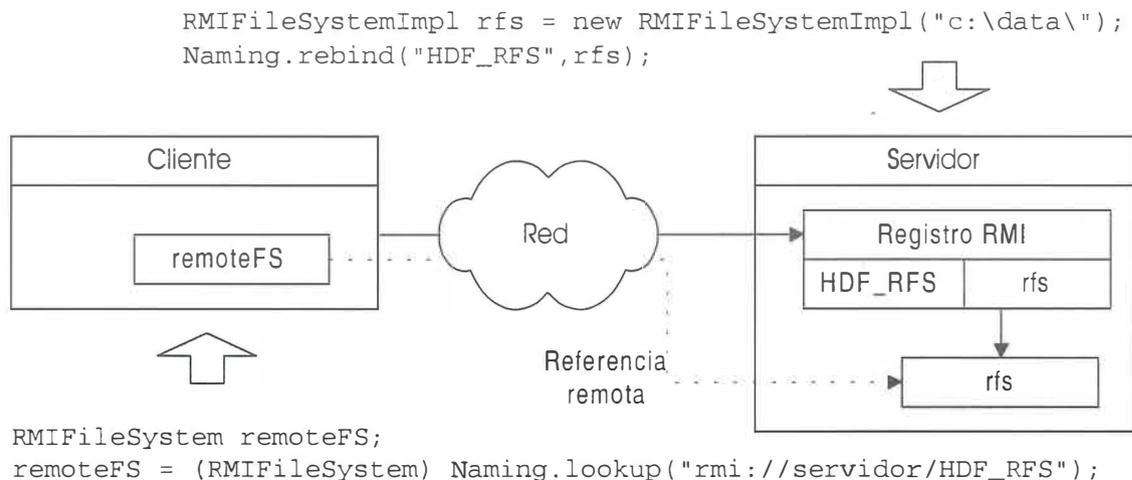


Figura 20. Sistema de nombramiento de RMI

Para que los clientes puedan invocar los métodos de un objeto remoto es necesario que primero obtengan una referencia a él. Esta referencia se encuentra registrada en el servicio de nombramiento de RMI (RMI Naming Service) o registro de RMI. Los servicios de nombramiento de RMI son accedidos a través de la clase `Naming`. En esta clase se definen los métodos necesarios para actualizar el contenido del registro de RMI. El método `bind` de la clase `Naming` crea un mapeo de un objeto remoto con un nombre que identifique a este objeto. El método `rebind` actualiza el mapeo entre un nombre y un objeto remoto en el registro de RMI. La clase `Naming` permite recuperar y definir un objeto

remoto del registro de RMI utilizando la sintaxis de los URL (Uniform Resource Locator), que en el caso de RMI son de la siguiente forma: `rmi://host:puerto/nombre`. Host es el nombre del servidor en donde está activo el registro de RMI, puerto es el número de puerto del servicio de registro de RMI (1099 es el valor por default) y nombre es el nombre que tiene asociado el objeto remoto en el registro. Un ejemplo de esto es `rmi://thechild.cicese.mx/2000/HDF_RFS`.

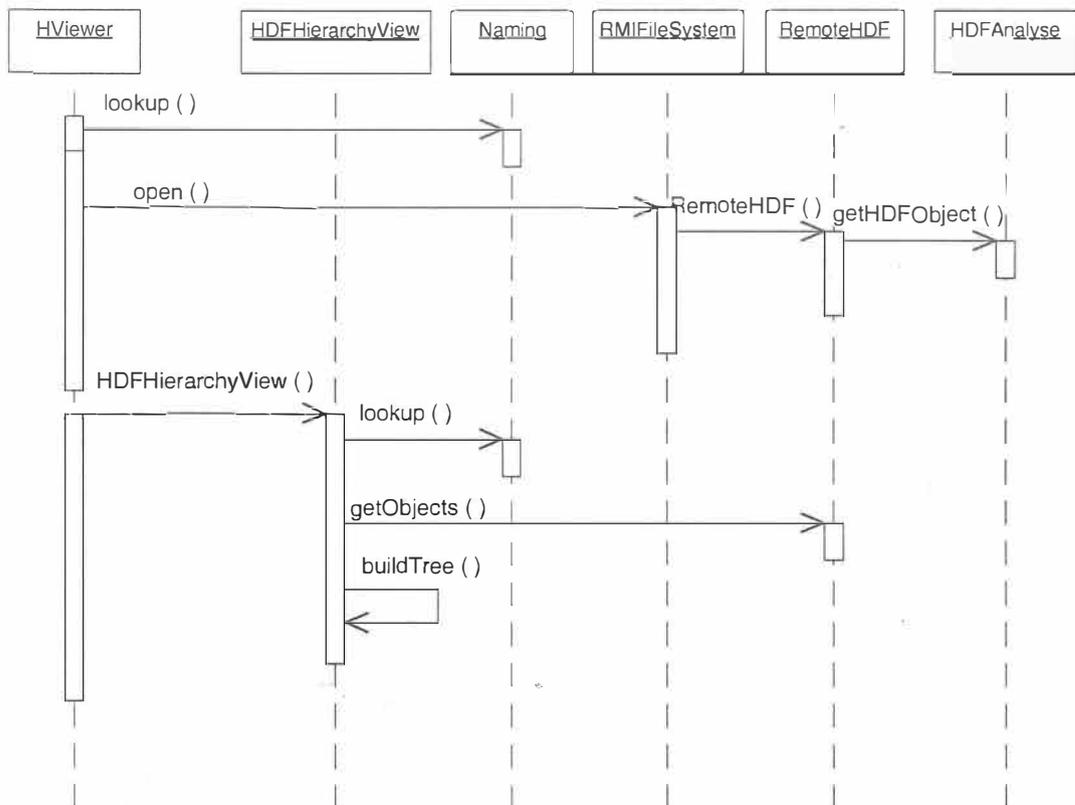


Figura 21. Diagrama de secuencia de la recuperación de objetos de datos en un archivo HDF remoto.

En la Figura 21 se presenta el mecanismo de recuperación del contenido de un archivo HDF del servidor. `RMIFileSystem` y `RemoteHDF` son los objetos remotos creados por `RemoteHDFServer` (servidor de datos). El proceso es similar al descrito en la Figura 13, la diferencia es que en este caso los objetos en el sistema cliente (`HViewer` y `HDFHierarchyView`) aprovechan los servicios que provee el sistema `RemoteHDFServer` a

través de los objetos remotos `RMIFileSystem` para el acceso remoto al directorio en el que se encuentran los archivos compartidos de datos y para seleccionar un archivo de ese directorio; y de `RemoteHDF` para recuperar los datos almacenados en un archivo en particular. El proceso se inicia cuando un usuario decide abrir un archivo ubicado en el banco de datos compartido, ya que se seleccionó el archivo, se hace la petición al servidor de archivos (`RMIFileSystem`) para que abra el archivo. Este verifica si el archivo no ha sido abierto con anterioridad, para esto se mantiene un registro de aquellos archivos que se encuentran abiertos o activos que se consulta cada vez que se hace una petición de abrir un archivo. Si el archivo no está abierto, es necesario crear un nuevo objeto `RemoteHDF` que se encargue de responder a las solicitudes de recuperación de datos almacenados en ese archivo en particular. Este objeto es el que interactúa con las librerías de HDF para tener acceso a los datos almacenados en los archivos y enviar los datos a los clientes. Una vez creado el objeto, obtiene la jerarquía de datos del archivo que se envía de regreso al cliente como una cola con la que se crea un árbol jerárquico de datos en los clientes.

Para ejemplificar el mecanismo de recuperación de datos de un archivo remoto, en la Figura 22 se describe la recuperación de objetos `Vdata`. Este tipo de dato que consiste en registros de datos cuyos valores se almacenan en campos de longitud fija, incluyen también información adicional como su nombre, la clase de `vdata` a la que pertenece y los nombres de los campos. Comúnmente estos datos se recuperan en forma de una tabla. Con la ayuda del árbol jerárquico de datos (`HDFHierarchyView`) se selecciona el objeto `Vdata`. Al seleccionarlo se activa la invocación al método remoto `getVdata` del objeto `RemoteHDF`

que está a cargo del archivo en el que se encuentra el objeto Vdata. En respuesta a la invocación del método `getVdata` se crea el objeto `VDataModelFactory` al que se le delega la responsabilidad de crear, con los datos almacenados en el archivo, la tabla de registros Vdata (`VdataTableModel`). Esta tabla se envía al cliente como valor de retorno para crear `VdataTableFrame`, componente de despliegue para la tabla (Figura 23).

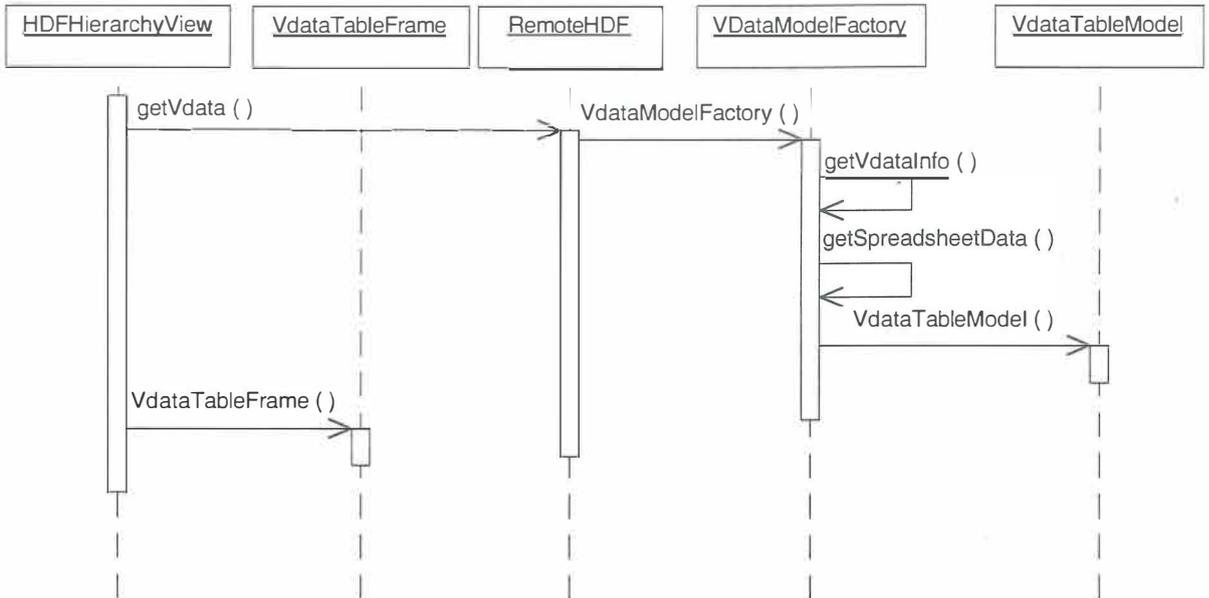


Figura 22. Diagrama de secuencia de la recuperación de objetos Vdata de un archivo HDF.

REFLAT	REFLON	DSPLAT	DSPLON	DELTAX	DELTAY	ROTANGLE
81.1428	-125.2682	81.1558	-125.3022	1.3	0.8	-1.0
81.188	-125.2176	81.201	-125.2516	1.3	0.8	-1.0
81.2333	-125.1664	81.2463	-125.2006	1.3	0.8	-2.0
81.0897	-125.0263	81.1028	-125.0597	1.3	0.8	-1.0
81.1349	-124.9748	81.1462	-125.0103	1.1	0.8	0.0
81.1801	-124.9227	81.1914	-124.9583	1.1	0.8	0.0
81.2253	-124.87	81.2365	-124.9058	1.1	0.8	0.0
81.2705	-124.8169	81.2817	-124.8528	1.1	0.8	0.0
81.3157	-124.7631	81.3287	-124.797	1.3	0.8	-1.0
81.0817	-124.7349	81.0947	-124.7678	1.3	0.8	-1.0
81.1268	-124.6819	81.1381	-124.717	1.1	0.8	0.0
81.172	-124.6283	81.1832	-124.6636	1.1	0.8	0.0

Figura 23. Despliegue de una tabla de los registros Vdata de un archivo HDF.

### IV.2.3. Interface de usuario del sistema cliente.

La Figura 24 representa la jerarquía de objetos de datos de un archivo remoto HDF en los sistemas clientes. Las diferencias existentes con la forma en que se hacía esto en el prototipo anterior (Figura 14), son que en primer lugar se cambió la clase encargada de ésta operación HDFTreeView por la clase HDFHierachyView, la cual implementa la misma funcionalidad que la clase anterior, pero auxiliándose en los objetos remotos y no en la interface JHI. Adicionalmente se modificó ligeramente la representación de la raíz del árbol jerárquico de datos de tal manera que ésta indique en que servidor de datos se encuentra ubicado el archivo HDF.

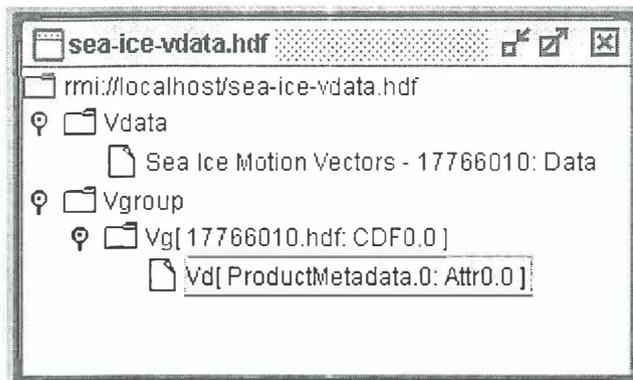


Figura 24. Representación de la jerarquía de objetos de datos de un archivo HDF remoto.

La interface de usuario del sistema cliente es muy parecida a la interface que presenta el primer prototipo (Figura 15) del sistema. La Figura 25 presenta la interface de usuario del sistema cliente.

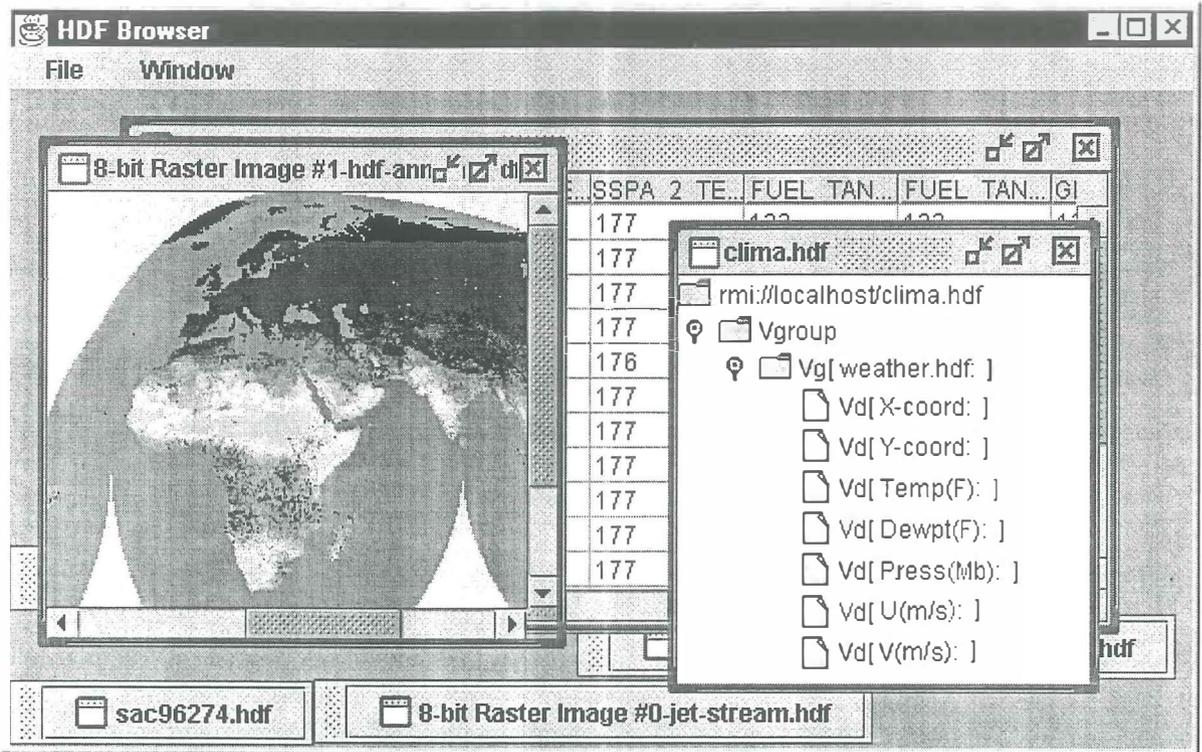


Figura 25. Interface de usuario del sistema cliente.

La interface conserva el modelo MDI presentado en el prototipo anterior. Una de las características adicionales de este sistema cliente es que los archivos se pueden acceder de distintos servidores al mismo tiempo. Con ello es posible el acceso simultáneo a distintos bancos de datos que estén corriendo el sistema RemoteHDFServer sin tener que cortar la conexión de los archivos abiertos en un banco de datos para conectarse a otro.

### IV.3. Desarrollo del tercer prototipo.

En este prototipo se extendió la funcionalidad del prototipo con el objetivo de crear una herramienta cooperativa para la visualización de datos científicos que permita a investigadores en distintas ubicaciones visualizar simultáneamente archivos almacenados en un banco de datos, formando un ambiente compartido para el análisis de los datos y la facilitación del diálogo, elemento esencial para la colaboración. La Figura 26 presenta el

modelo de un ambiente de colaboración que provee un espacio compartido como un auxiliar para comunicación.



Figura 26. Modelo de un ambiente de colaboración.

El acceso compartido a información permite que la representación simbólica se haga más concreta. Stefik *et al.* [1987] utilizan al pizarrón como metáfora de un espacio compartido de información. De aquí surge la idea de las interfaces de tipo WYSIWIS (What You See Is What I See: lo que ves es lo que veo). Este tipo de interfaces provee la presentación de imágenes consistentes de información compartida a todos los participantes de una colaboración.

En las aplicaciones típicas de visualización el usuario se encuentra aislado con los datos que está analizando. En estas y otras aplicaciones es común que cuando alguien observa algo que le parece interesante, llame a otros compañeros o colaboradores para que se acerquen a observar lo que se presenta en el monitor (i.e. una gráfica, el código de un programa o una página del web), lo que generalmente se acompaña de algún diálogo o discusión relativo a lo que se está observando. De ésta manera la pantalla de la computadora sirve como espacio compartido de información. Sin embargo, éste tipo de interacción no se puede efectuar cuando los colaboradores se encuentran en otro edificio o en otra ciudad. Surge así la necesidad de proveer de un ambiente que facilite este tipo de interacciones en el

proceso de visualización de datos científicos a distancia. La Figura 27 presenta un modelo de éste ambiente.

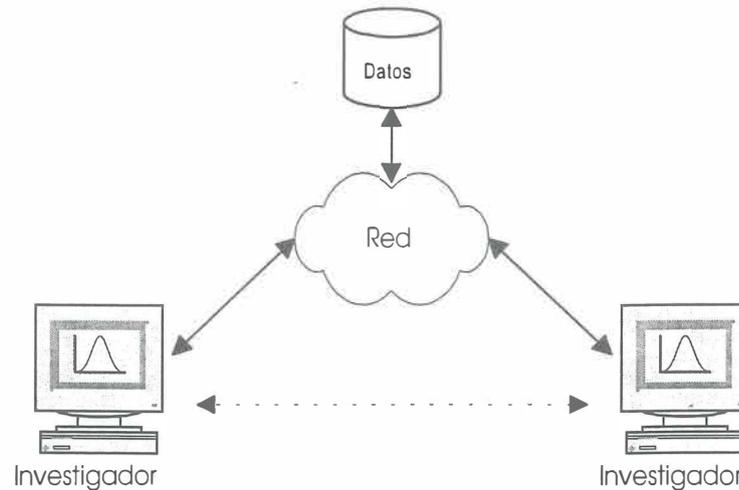


Figura 27. Modelo de un ambiente colaborativo de visualización.

#### IV.3.1. Caracterización del ambiente colaborativo de visualización.

Soriano [1996] definió una caracterización del espacio de diseño de sistemas colaborativos (CEDSC) el cual incluye requerimientos funcionales y decisiones de diseño (tipo de interacción, participación, visualización, notificación, coordinación, distribución y confidencialidad) que proveen una visión global de los aspectos que se deben de considerar en el desarrollo de un sistema colaborativo. A partir de esta caracterización, Zapata [1997] creó una metodología para el diseño de arquitecturas de sistemas colaborativos síncronos. Esta metodología identifica las relaciones que existen entre los requerimientos funcionales y las decisiones de diseño para utilizarlas como reglas de diseño que muestren distintas alternativas de implementación del sistema.

La metodología requiere como primer paso la descripción general de la funcionalidad del sistema. En el caso particular del sistema que se desarrolló, se tienen como

requerimientos principales crear un ambiente colaborativo para la visualización de datos que permita la interacción de un grupo pequeño (un máximo de 6 integrantes) de investigadores en distintas ubicaciones durante el proceso de análisis de datos científicos almacenados en un banco de datos. Para esto se requiere que los usuarios tengan acceso remoto a los datos, que los mismos datos se desplieguen en forma simultánea a todos los participantes y que se proporcionen áreas de trabajo común que auxilien la interacción y un medio de comunicación para que los participantes puedan dialogar.

El siguiente paso que establece la metodología consiste en describir en forma breve los requerimientos funcionales de carácter social más importantes como el tipo interacción y participación. Este sistema debe facilitar el establecimiento de sesiones de colaboración científica entre un grupo de investigadores. La interacción será síncrona y distribuida. Debido a la naturaleza de la tarea que se está auxiliando con éste ambiente, se tiene que las sesiones de trabajo requieren de cierta formalidad durante su desarrollo, los participantes tienen un mismo nivel jerárquico y debe de prevalecer la democracia a lo largo de una sesión de trabajo. Para la realización de una sesión de trabajo será necesario hacer un convocatoria con anterioridad a la sesión. Durante la sesión se deberá de seguir una agenda que sea flexible en lo referente a puntos a tratar. Debe ser posible identificar las aportaciones de cada participantes y la realización de las tareas se harán en forma conjunta. Cada uno de los participantes tendrá dedicado cierto tiempo para hacer sus aportaciones.

Con la información anterior, se procede a llenar un formato de caracterización del espacio de diseño del sistema a desarrollar. Posteriormente se realiza un revisión del formato

en el cual se revisa cada uno de los elementos de la taxonomía y se aplican las reglas definidas por la metodología con lo cual se obtiene un nuevo formato. La Figura 28 muestra el formato de caracterización que se obtuvo con éste proceso. En la revisión, se contó con un total de 16 diferencias en los elementos de cada uno de los formatos. Algunas de estas diferencias se debieron a que para el llenado del primer formato se mal interpretaron algunos de los elementos del espacio de diseño. Otras diferencias se debieron a la aplicación de las reglas sobre elementos relacionados.

System : Colaboratorio para Estudio del Fenómeno "El Niño" Junio 98

**Location:**  
 co-located  distributed  predictable  unpredictable  mixed mode

**Time:**  
 synchronous  asynchronous  predictable  unpredictable

**Type of communication:**  formal  informal

**Social structure of the session:**  democratic  stratified

**Social status of participants:**  same  hierarchical

**Invocation to participate:**  scheduled  spontaneous

**Interaction flow of the session:**  structured  unstructured  
 inflexible  flexible

**Authorship of ideas:**  anonymous  identified  role

**Styles of accomplishing larger task:**  joint  by subgroups

**Access to a public area:**  turn-taking  simultaneous

**Type of contribution:**  independent  reflective  by consensus  partitioned  recorded

**Method to establish a session:**  predefined  dynamic

**Joining a session:**  on time  late comers  intervals

**Leaving a session:**  before it concludes  on finishing

**Policy to give access to late comers:**  
 accepted  consensus  moderator  vote  criterion

**Group size:**  2 persons  small (3-6)  large (+6)

**List of attendees:**  not specified  specified  conditioned

**Forms of accomplishing activities:**  joint  parallel

**Communication mechanisms:**  text  sketch  audio  video

**Annotations:**  static  dynamic

**Endurance of data elements:**  persistent  not persistent

**Period of interaction:**  task  environment  propagation  perceptive

**Data representation:**  identical  different

**Actions and functions:**  listing  drawing  gesturing  
 storing information  expressing ideas  mediating interaction

**Mult user interface, WYSIWIS:**  
 strict  relaxed  not available  
 space  time  population  congruence

**Type of windows:**  private  public  shared

**Visual cues to identify windows:**  
 different  title  icon  color  to organize independently  
 no visual cue

**Screen workspace management:**  
 independent iconification  miniature scale  change window size  select which are full size

**Data transmission to maintain simultaneous views:**  
 broadcast each update  after reasonable time

**Maintaining user displays:**  updated  not updated

**Window representation change when active:**  
 deiconified  update the icon  indication bar  flickering window

**Cursor(s) used:**  particular  multiple  common  shared  public

**Display of the cursors:**  all  local  telepointer

**Selection of menu items:**  private  widespread  semantic cursors

**Collaboration/workspace awareness:**  
 transparent  awareness  
 presence  location  changes  actions  objects  viewing area  activity level  abilities  intentions  sphere of influence  expectation

**Visualization of spatial location:**  
 multi user scrollbar  gestalt views  radar  portrait radar  slave view  miniature view  WYSIWID windows  fisheye view  video images  none

**Graphic representation of participants:**  
 none  icon images  named cursor  video images  users list  status line  Other :

**Availability of participants:**  
 accessible  interrupted  not interrupted

**Premeeting process:**  none  goal setting  signing participants  collecting material

**Support for particular styles of meetings:**  casual interaction  facilitated group support

**Registration style for applications:**  same style different tools  different styles same tools

**Registration schemes:**  
 open door  facilitator controlled  rooms based  communities with shared interest  Other:

**Dimensions for control floor policies:**  
 number of floors 1 number of people 1 how the floor is passed \_\_\_\_\_

**Floor control mechanisms:**  
 free floor  pre-emptive  round robin  moderator controlled  random  FIFO with explicit release  explicit release  pause detection

**Factors for selecting a concurrency control method:**  
 human considerations  overall network and processor latency  resources used  number of network transactions  memory requirements

**Floor policy to request the floor and how it is granted (request/grant):**  
 implicit/implicit  explicit/implicit  explicit/explicit  no floor

**Visual representation of current floor holder:**  
 panel  title  cursor  none

**Concurrency control mechanism:**  
 no control  locking  serialization  Other:

**Concurrency control in architectures:**  
 centralized  replicated

**Data architecture:**  centralized  replicated

**Application architecture:**  centralized  replicated  hybrid

**Communication channels:**  LAN  WAN  audio/video

**Programming toolkits:**  used  not used

Figura 28. Formato de caracterización del espacio de diseño del ambiente colaborativo de visualización.

### IV.3.2. Descripción del ambiente de visualización colaborativo.

La Figura 29 muestra la interface de usuario del ambiente colaborativo de visualización. El tipo de interface que presenta el ambiente es del tipo WYSIWIS relajado con lo que se provee mayor flexibilidad en las interacciones. Se conserva del modelo MDI utilizado en los prototipos anteriores y se agrega la utilización de ventanas de tipo público, lo que implica que todos los participantes en una sesión tendrán a la vista el mismo conjunto de ventanas e información así como las mismas herramientas de trabajo. Adicionalmente se permite que el usuario trabaje en forma individual, sin necesidad de que participe en una sesión de trabajo con otros colaboradores para tener acceso al banco de datos.

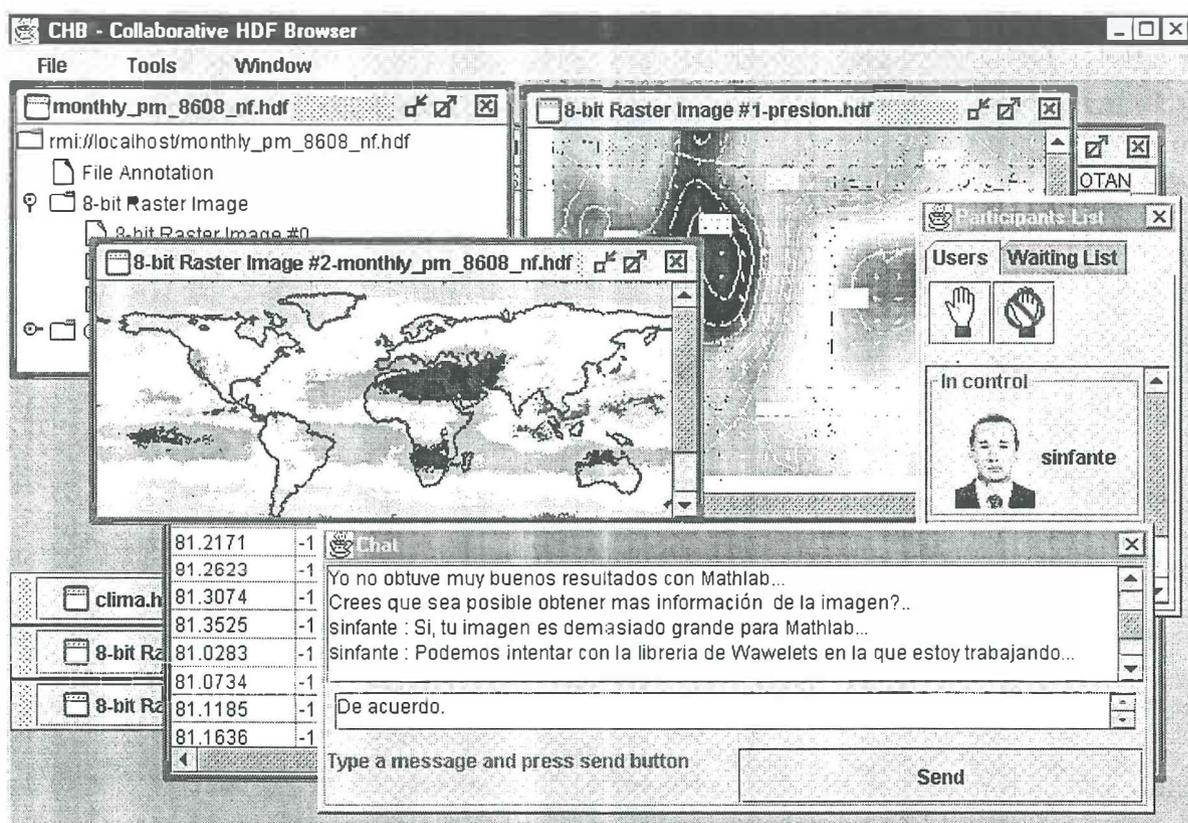


Figura 29. Interface de usuario del ambiente colaborativo de visualización.

El ambiente está constituido por varios elementos. Un visualizador colaborativo de archivos y un conjunto de herramientas de apoyo a la colaboración. Durante una sesión los participantes tienen acceso a un conjunto de herramientas para facilitar la interacción con otros participantes. Primeramente se provee de un medio de comunicación textual (chat) con el cual los participantes puedan intercambiar mensajes para sostener conversaciones acerca de la tarea que se está realizando (Figura 30). Se seleccionó éste medio de comunicación porque para su funcionamiento, se requiere de un mínimo de recursos en comparación de otros medios como el audio o vídeo. En la actualidad existen algunos productos, i.e. CuSeeme y NetMeeting ambos disponibles gratuitamente, que ofrecen éste servicio y que se pueden utilizar en forma conjunta con este ambiente.

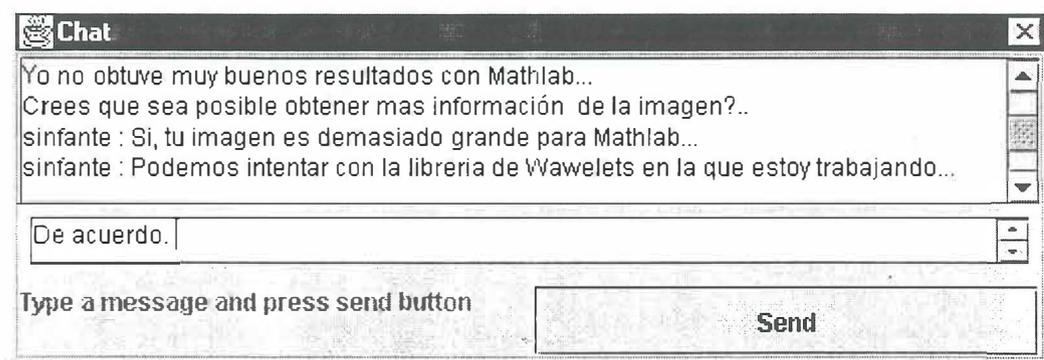


Figura 30. Chat para el intercambio de mensajes textuales.

Otra herramienta que proporciona el ambiente es un pizarrón compartido sobre el cual es posible hacer anotaciones gráficas sobre imágenes para resaltar elementos sobresalientes en la imagen, que sirve también como medio de comunicación gráfica para expresar ideas. La Figura 31 muestra el empleo de esta herramienta para hacer anotaciones sobre una imagen extraída de un archivo HDF. No está limitada a imágenes extraídas de un archivo HDF almacenado en el banco de datos, sino que también permite al usuario compartir

archivos de imágenes en formato GIF o JPEG almacenados en su propia estación de trabajo. Con la finalidad de identificar las aportaciones que hace cada participante cuando hace alguna anotación, cada participante puede escoger un color que identifique los trazos que realice en el pizarrón.

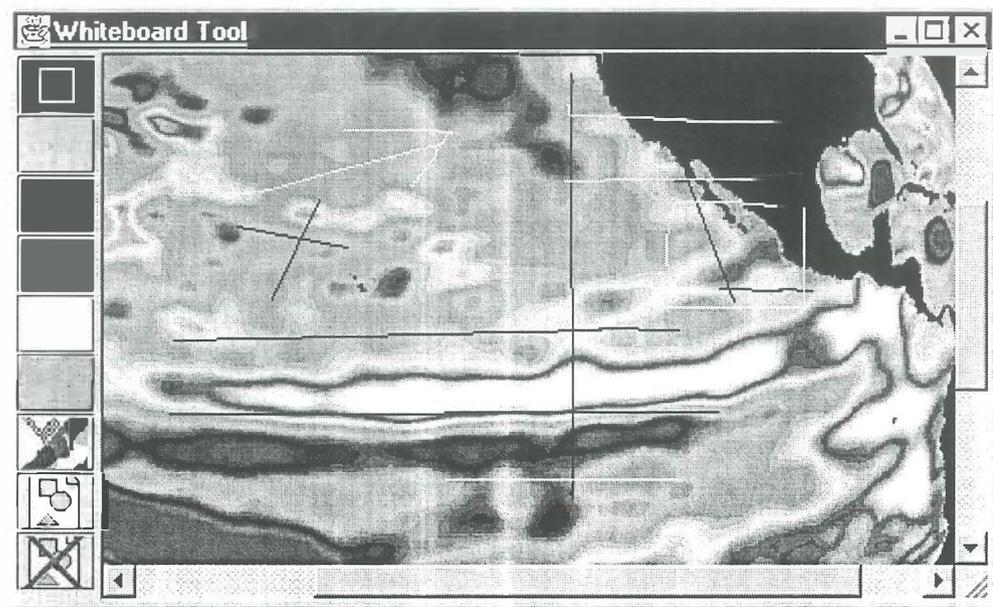


Figura 31. Pizarrón compartido para realizar anotaciones gráficas.

En la mayoría de las sesiones de trabajo es necesario utilizar información almacenada en diferentes tipos de documentos (i.e. artículos, reportes técnicos, agendas, minutas, entre otros) para exponerla al resto de los participantes. Para lo anterior, se proporciona a los usuarios una herramienta que permite visualizar documentos hipertexto de manera compartida. Esta herramienta incluye la funcionalidad básica de un browser de WWW y agrega la sincronización de despliegue como característica propia. Cuando un usuario sigue una de las ligas del documento, los visualizadores de documentos del resto de los participantes siguen automáticamente esa misma liga para mantener la sincronización de la información desplegada. La Figura 32 muestra la herramienta.

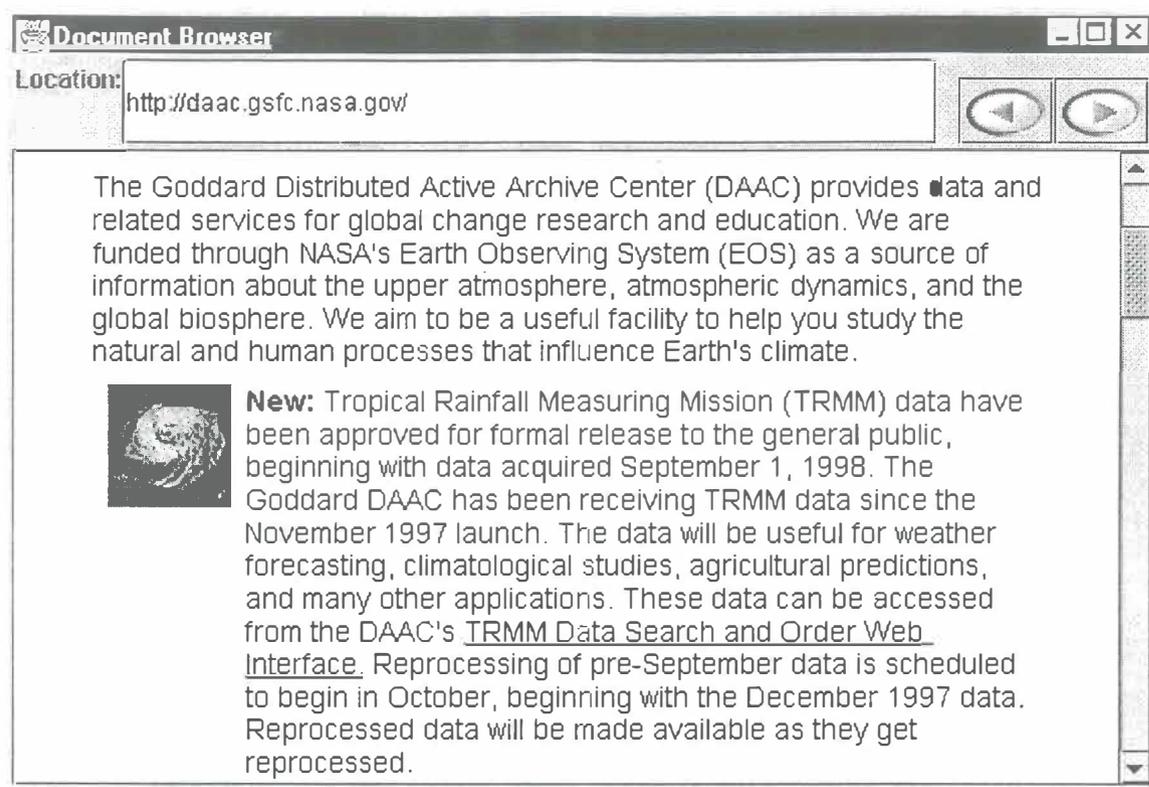


Figura 32. Visualizador de documentos hipertexto.

Todas las herramientas anteriores están integradas al visualizador colaborativo de archivos HDF, elemento principal de éste ambiente. El visualizador tiene la función de mantener la sincronización del despliegue de los datos recuperados del bancos de datos entre los participantes de una sesión. Cualquier acción relacionada con el manejo de los datos realizada por alguno de los participantes, como la apertura de un archivo o la recuperación de algún objeto del archivo, se difunde al resto de los visualizadores conectados en la sesión para que actualicen el despliegue de información. De esta manera se mantiene la consistencia de la información visualizada entre todos los participantes [Stefik *et al.*, 1997].

Otros mecanismos que ayudan a mantener la consistencia de los despliegues y a coordinar la interacción de los participantes son los mecanismos de control de piso. Estos

mecanismos regulan el acceso a los elementos de un espacio compartido y la intervención de los participantes durante una sesión [Roseman, 1992]. Cuando en una reunión un participante tiene su oportunidad para intervenir, se dice que tiene el control de piso. Los mecanismos de control de turno incluyen distintas políticas para el control de piso: libre, aleatorio, demanda, uso de colas con liberación explícita, uso de cola circular, moderador o detección de pausa [Soriano, 1996].

El ambiente colaborativo de visualización utiliza un cola o lista de espera con liberación explícita como control de piso. En esta política se tiene un lista de participantes de tipo primero en entrar, primero en salir. Cuando el participante que tiene el control de piso termina su intervención, explícitamente cede el control de piso y éste se asigna automáticamente al primer participante en la lista. Desde el punto de vista de los participantes éste mecanismo sería por petición explícita con asignación implícita. Ya que se requiere que el usuario explícitamente solicite el control de piso, ingresando de ésta manera a la cola de espera y el control se le otorga en forma implícita al llegar al inicio de la fila. La Figura 33 muestra como se representan los participantes de una sesión. Además de servir como referencia a los participantes acerca de con quien están colaborando, la lista de participantes tiene las funciones de notificar a los participantes quien tiene el control de piso asignado, así como para que los participantes soliciten o liberen el control de piso y desplegar la lista de participante en espera del control de piso.

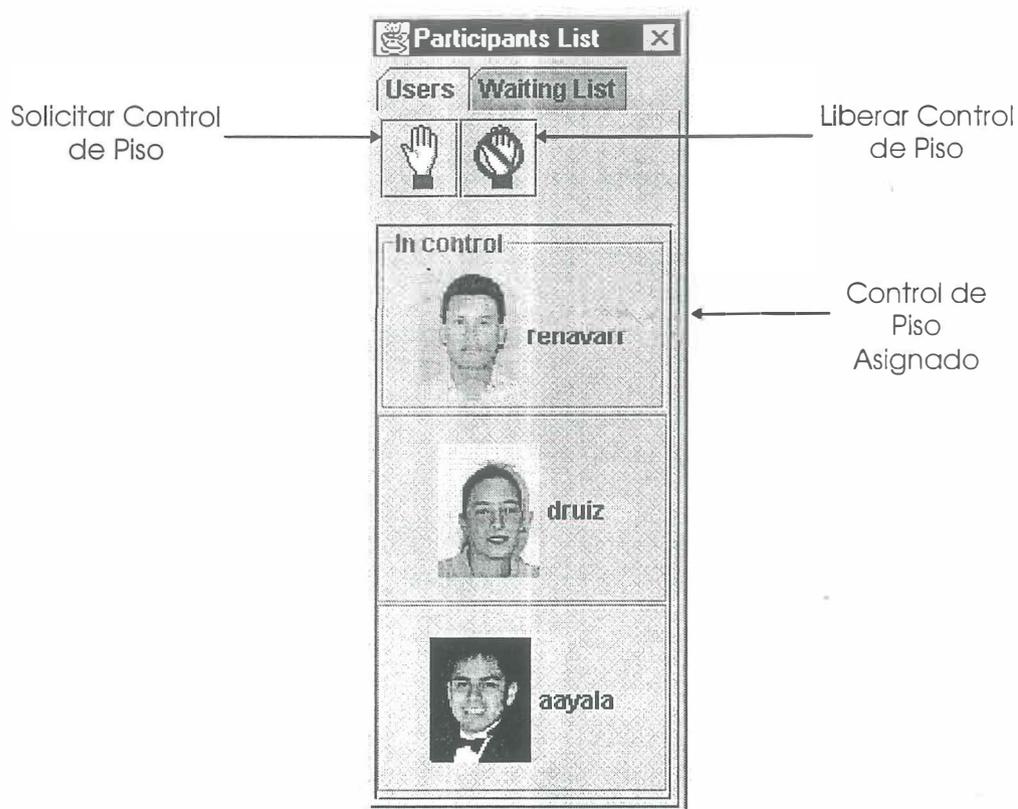


Figura 33 Lista de participantes en una reunión.

Como se indica en la Figura 33, la solicitud de control de piso se hace presionando el botón que presenta el icono de una mano levantada. Cuando un participante tiene asignado el control de piso, es el único que tiene el derecho para manejar los distintos elementos compartidos que se despliegan en la pantalla, abrir archivos, recuperar datos de un archivo, hacer anotaciones en el pizarrón y utilizar el visualizador de documentos. El chat es la única herramienta que puede ser utilizada sin necesidad de tener el control de piso asignado. Cuando el participante termina su intervención, éste libera el control pulsando el botón indicado en la Figura 33. Una vez que lo ha liberado, se asigna automáticamente al primer participante en la lista de espera.

### IV.3.3. Arquitectura del ambiente colaborativo de visualización.

Para la implementación de este sistema se extendió la funcionalidad del servidor de datos (`RemoteHDFServer`) de tal manera que se incluyeran los mecanismos para el manejo de las sesiones (i.e. permitir la conexión de varios participantes en una misma sesión) y para el manejo de los eventos que se generen en las sesiones (i.e. hacer una transmisión para notificar la ocurrencia de un evento a los sistemas clientes). En la Figura 34 se detallan las modificaciones realizadas.

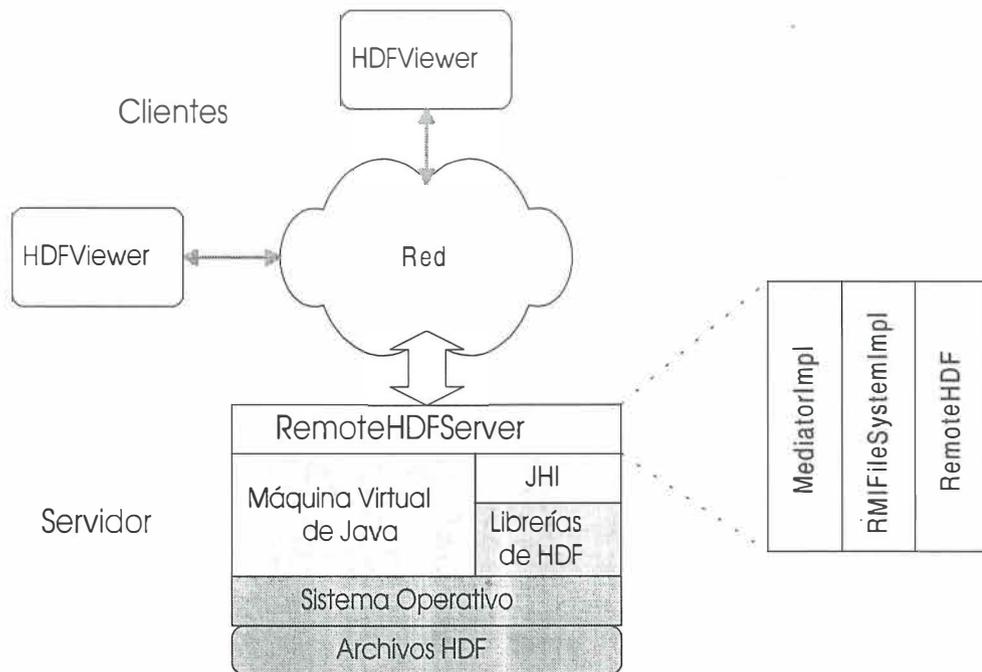


Figura 34. Arquitectura del ambiente colaborativo de visualización.

Para agregar la funcionalidad antes mencionada fue necesario incluir dos nuevos elementos a la arquitectura existente, un administrador de sesiones (`MediatorImpl`) en la parte del servidor y un manejador de eventos (`CollaboratorImpl`) en la parte del cliente, los cuales funcionan como objetos remotos (Figura 35). La unión de estos dos elementos proporcionan un mecanismo centralizado de comunicación entre los sistemas clientes que

intervienen en una sesión. Para mantener la sincronización del despliegue de la información en los clientes, es necesario que cuando en uno de los clientes se realice alguna acción que genere alteraciones en el contenido de la información desplegada (i.e. apertura de un nuevo archivo), éste evento se notifique al resto de los clientes para que actualicen su despliegue. Cualquier evento de éste tipo es enviado primeramente por los clientes al manejador de eventos del servidor (`MediatorImpl`), el cual se encarga de notificar a todos los clientes conectados en esa sesión.

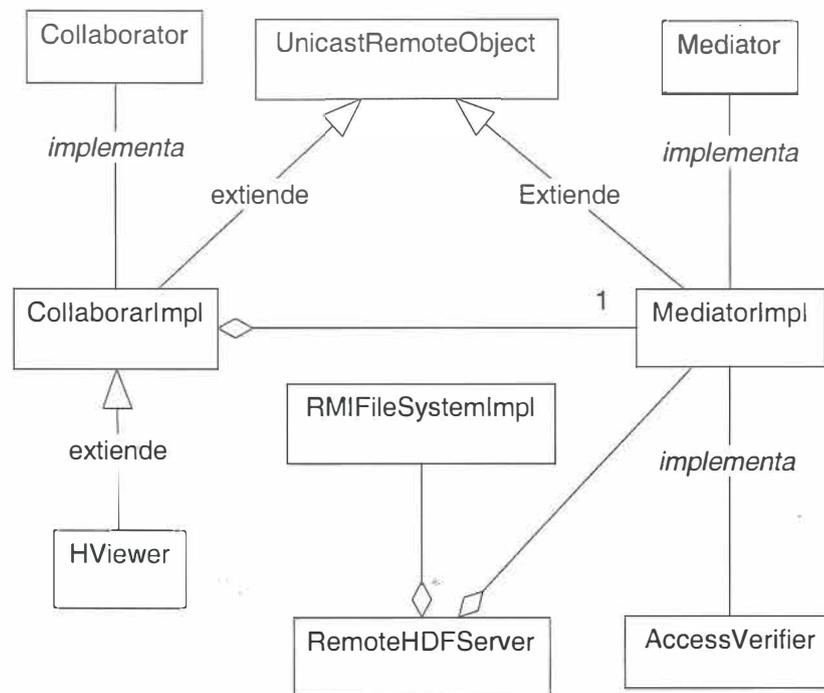


Figura 35. Diagrama de objetos del ambiente colaborativo de visualización.

La comunicación entre los clientes puede ser punto a punto, cuando un cliente requiere notificar algún evento o enviar cierta información a un cliente en particular, o bien multipunto cuando un cliente requiere comunicarse con todos los clientes conectados. Esta comunicación se efectúa por medio de mensajes especificados por un protocolo que define

el conjunto de mensajes válidos. La estructura de los mensajes y algunos ejemplos de mensajes predefinidos se presentan en la Tabla IV.

Tabla IV. Ejemplos de mensajes definidos por el protocolo de comunicación.

Etiqueta (String)	Argumentos	Descripción
NEW_USER	Identity	Se ha incorporado un participante a la sesión.
DROP_USER	Identity	Un participante ha abandonado la sesión.
NEW_HDF_FILE	String	Se ha abierto un archivo.
DISP_RIS8	HDFObjectNode	Obtener y desplegar imagen raster.
DISP_VDATA	HDFObjectNode	Obtener y desplegar registros VDATA.

Los mensajes están compuestos por una etiqueta que identifica al mensaje y los argumentos que incluyen cualquier información adicional que se requiera. Por ejemplo, el mensaje NEW\_USER (Tabla IV) notifica de la incorporación de un nuevo participante a una sesión. El argumento que acompaña al mensaje incluye información relacionada con identidad del nuevo usuario encapsulada en un objeto de tipo Identity.

El proceso completo de incorporación de un participante a una sesión se muestra en la Figura 36. Este inicia en el momento en que el participante ejecuta el programa del sistema cliente (HViewer). Al ejecutarse el programa cliente, este pide al participante su nombre de usuario (login) y su clave de acceso al ambiente de colaboración. Una vez que obtiene esos datos, el sistema cliente obtiene una referencia al manejador de sesiones a través del registro de RMI. Ya que se tiene localizado al manejador de sesiones, realiza la invocación del método remoto `checkAccess` para verificar que el participante esté registrado y tenga autorización para utilizar el ambiente. Si el participante está registrado, el sistema cliente se conectará a la sesión y se dará de alta en el administrador de sesión. Posteriormente se

procede a notificar al resto de los clientes que un nuevo participante se ha incorporado esto se hace mediante el llamado al método `broadcast` del administrador de sesiones el cual se encargara de hacer la notificación a los clientes para que actualicen su lista de participantes.

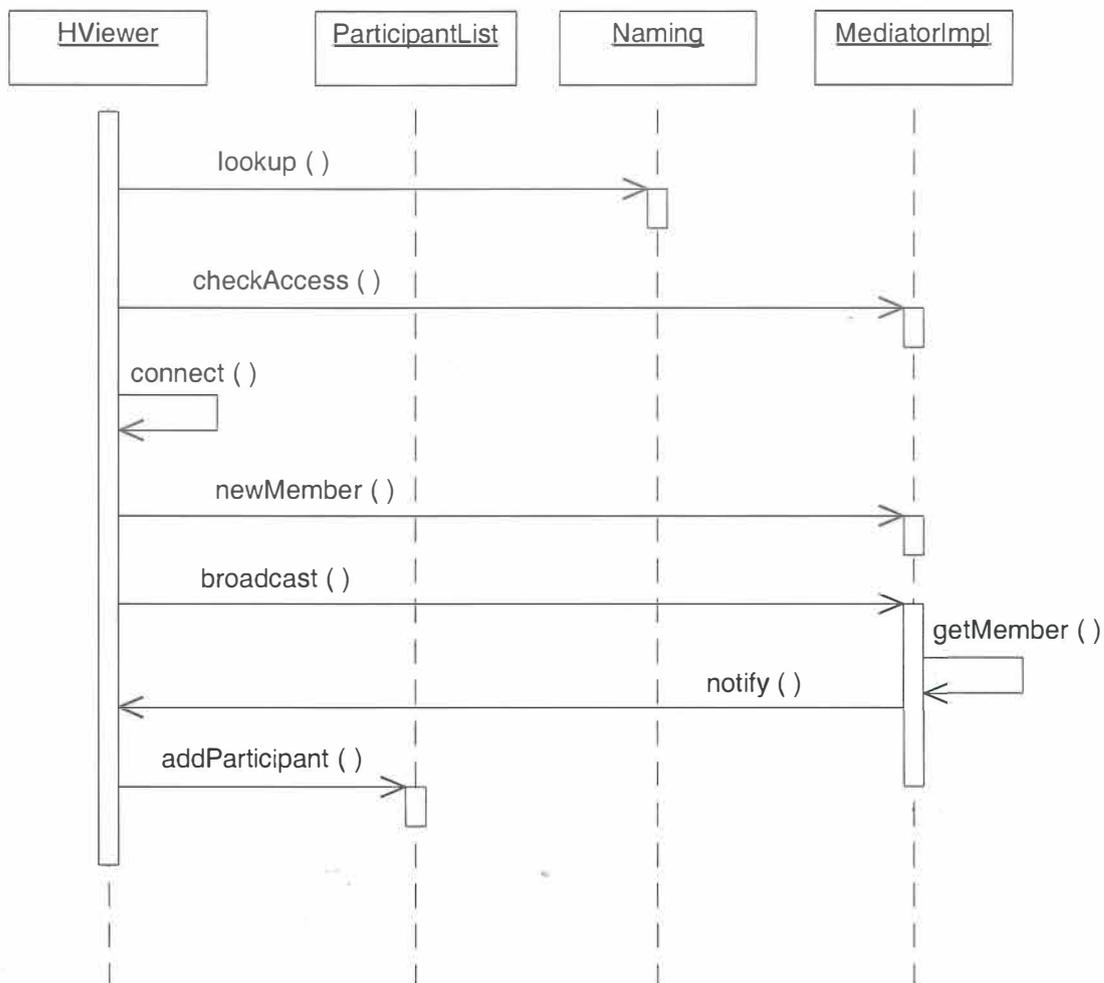


Figura 36. Diagrama de secuencia del proceso de registro de un nuevo participante.

## CAPÍTULO V EVALUACIÓN DEL AMBIENTE COLABORATIVO DE VISUALIZACIÓN

Para la evaluación del ambiente se realizaron pruebas de usabilidad, definida esta como la cualidad de un sistema que lo hace fácil de aprender, fácil de recordar, tolerante a los errores y subjetivamente satisfactorio [Pressman, 1997]. La usabilidad de los sistemas de software es uno de los principales objetivos dentro de la interacción humano-computadora. Desde el punto de vista de los usuarios, la usabilidad puede hacer la diferencia entre realizar una tarea con exactitud o no, completar una labor o dejarla incompleta, disfrutar el proceso o acabar frustrado.

La evaluación se realizó utilizando el protocolo de *expresión verbal del pensamiento* (thinking aloud protocol), que es una técnica para realizar pruebas de usabilidad [Nielsen, 1994]. Durante el desarrollo de la prueba los participantes realizan una tarea como parte de un escenario de uso, se pide a los participantes que mientras interactúan con el sistema expresen verbalmente sus pensamientos, emociones y opiniones. Esta técnica permite comprender de que manera el usuario interactúa con la interface de usuario y que consideraciones tiene el usuario en mente cuando utiliza la interface. Otra ventaja adicional es que la terminología que el usuario emplea para expresar una idea o función, puede incorporarse al diseño del sistema.

### V.1. Escenario de ejecución de las pruebas.

Las pruebas se realizaron en el marco del Coloquio Sobre Oceanografía del Pacífico Oriental (realizado en las instalaciones del CICESE, en septiembre de 1998) y con la ayuda de participantes voluntarios escogidos entre los asistentes al coloquio. Ninguno de los

voluntarios contaba con experiencia en la utilización de sistemas colaborativos. Sin embargo, la mayoría contaba con experiencia en el manejo de aplicaciones de software para el análisis y visualización de datos científicos. Para efectos de la ejecución de las pruebas se conectaron dos computadoras a la red del edificio de la División de Oceanología con un prototipo totalmente funcional del sistema cliente del ambiente colaborativo de visualización instalado. El servidor de archivos (`RemoteHDFServer`) se instaló en una máquina ubicada en el edificio de la División de Física Aplicada. En esta máquina se tenían almacenados cerca de 40 archivos HDF la mayoría con información atmosférica (i.e. niveles de temperatura, humedad, ozono y radiación) a nivel global.

Se realizaron tres sesiones de 10 minutos de duración en las cuales se involucró a seis voluntarios (dos en cada una de las sesiones) que fungieron como usuarios del sistema. De manera previa a cada sesión se presentó a los voluntarios una reseña muy general acerca de la funcionalidad del sistema a evaluar y de la mecánica de la sesión para proceder después a la ejecución de las pruebas. Durante la sesión la técnica de expresión verbal del pensamiento se complementó con preguntas directas acerca de los distintos elementos de la interface de usuario y funcionalidad del sistema (i.e. ¿Qué es lo que haría usted normalmente con estos datos? ¿Cómo organizaría las ventanas? ¿Son correctos los términos que se emplean en el sistema?). Se efectuaron también sesiones más breves (5 minutos), en las cuales sólo participaba un voluntario accedendo a los archivos almacenados en un servidor. En estas sesiones se dejaba al voluntario explorar libremente el sistema, ofreciéndole ocasionalmente algunas indicaciones sobre la funcionalidad del mismo.

## V.2. Resultados obtenidos de las pruebas.

En cuanto al uso de las herramientas de colaboración proporcionadas por el sistema, el pizarrón compartido fue la herramienta que se utilizó más frecuentemente. El chat se usó con menos frecuencia. Esto se debe a que los voluntarios durante la sesión estaban ubicados en un mismo lugar, por lo que se dio preferencia a la comunicación verbal directa. Esta misma, cercanía y el número reducido de los participantes provocó en ellos cierta incomodidad en el manejo de los mecanismos de control de piso, el cual parecía limitar la fluidez de la interacción.

La interface del sistema les pareció sencilla. Sin embargo, se comentó que las ventanas aparecían demasiado apiladas y que la ventana del chat y la lista de usuarios ocupaban demasiado espacio, por lo que se agregó la opción de poder ocultar las ventanas. Se modificaron también los títulos de algunas ventanas ya que resultaron confusos. Al pizarrón compartido se le agregó la opción de borrar solo las anotaciones realizadas, ya que originalmente sólo se tenía la opción de borrar el contenido del pizarrón completo.

Lo que llamó más la atención a los voluntarios fue la sencillez con la que se puede acceder a los archivos remotos. Se expresó que el mecanismo utilizado permite ahorrar tiempo en la identificación y selección de los datos de utilidad, ya que esta es una tarea que requiere de algo de tiempo. Si bien es cierto que la mayoría de los bancos de datos tienen su información catalogada, frecuentemente éstos catálogos no se encuentran actualizados. Aun cuando los catálogos están actualizados, primero se tiene que recurrir a estos y después acceder al banco de datos para recuperar los archivos de datos y verificar que en realidad los

archivos contienen los datos que ellos requieren. El mecanismo presentado por el ambiente les permite consultar directamente el contenido de los archivos sin hacer la transferencia del archivo completo.

## CAPÍTULO VI CONCLUSIONES Y TRABAJO FUTURO.

Con base en el concepto de Colaboratorios, en esta tesis se desarrolló un ambiente colaborativo para la visualización y análisis remoto de datos científicos. El cual está formado por interfaces de comunicación que proveen el acceso remoto, a un banco de datos científicos y herramientas de software, para facilitar la colaboración entre los investigadores del fenómeno El Niño en México.

### VI.1. Aportaciones.

Las contribuciones de este trabajo son:

- Un colaboratorio para estudios del fenómeno El Niño, el cual puede ponerse en operación con la participación de distintas instituciones (i.e. CICESE, CIBNOR, CICIMAR, IIO-UABC). Se consideró también la necesidad de implementar programas a nivel nacional que fomenten tanto la colaboración como el intercambio de datos entre instituciones de investigación financiadas con fondos públicos.
- Un marco de referencia que sirva de guía para la creación de otros colaboratorios. En este trabajo se presenta un modelo de desarrollo sencillo en el cual se combinan exitosamente técnicas bien comprobadas de la ingeniería de software (i.e. desarrollo de prototipos) con nuevas metodologías desarrollo de sistemas de groupware (i.e. caracterización del espacio de diseño de sistemas colaborativos). Adicionalmente, se introducen nuevas tecnologías para la construcción de sistemas distribuidos.

- Una arquitectura escalable de software para el acceso remoto, interactivo y simultáneo a múltiples bancos de datos científicos.
- Una herramienta colaborativa síncrona para la visualización y análisis de datos científicos que provee un medio conveniente para compartir datos con colaboradores.

## VI.2. Trabajo futuro.

- Incluir en el ambiente colaborativo de visualización facilidades para establecer interacciones casuales [Isaacs *et al.*, 1996]. Cuando se buscan colaboradores, los investigadores usualmente tratan de identificar colaboradores con intereses comunes. Este proceso requiere de contactos breves y espontáneos entre los colaboradores potenciales que se dificultan con la separación geográfica. El proveer los medios para propiciar estos encuentros sería de gran utilidad para aminorar ese problema.
- Extender la funcionalidad del ambiente para no sólo proveer los medios para compartir datos científicos, sino también material literario como artículos científicos o reportes técnicos. Para esto se podría implementar una biblioteca digital que de igual manera como se tienen bancos de datos científicos, sirva de repositorio compartido de información documental.
- Implementar una máquina de búsqueda que facilite el proceso de identificación y recuperación de los datos almacenados en el repositorio. La mayoría de los formatos de archivos para datos científicos incluyen metadatos o información adicional que

detalla su contenido. Esta información se podría aprovechar para determinar de manera automática el contenido de un archivo.

## LITERATURA CITADA

- Agarwal, D., W. E. Johnston, S. Loken, and B. Tierney. 1995. Tools for Building Virtual Laboratories. Proceedings of Computing in High Energy Physics, Rio de Janeiro, Brazil.
- Agarwal, D. A., S. R. Sachs and W.E. Johnston. 1998. The Reality of Collaboratories, *Computer Physics Communications* 110(1-3). 34-141 p.
- Baecker, R.M. 1993. Readings in Groupware and Computer Supported Cooperative Work. Morgan Kaufmann Publishers. San Mateo, CA. 873 pp.
- BioMOO. 1997. BioMOO, the biologists' virtual meeting place.  
<http://bioinformatics.weizmann.ac.il/BioMOO/>
- Brown, S. A., M. Folk, G. Groucher and R. Rew. 1993. Software for Portable Scientific Data Management. *Computers in Physics*, 7, 304-308 p.
- Boehm, B. 1988. A Spiral Model for Software Development and Enhancement. *IEEE Computer*. 21(5): 61-72 p.
- Brookhaven National Laboratory (BNL). 1998. Protein Data Bank Documentation.  
[http://www.pdb.bnl.gov/pdb-docs/doc\\_help.html](http://www.pdb.bnl.gov/pdb-docs/doc_help.html)
- Cerf, V.G. 1993. National Collaboratories: Applying Information Technology for Scientific Research. National Academy Press. First Edition. Washington, D.C. 118 pp.
- Chen, H., J. Martinez, T. Ng and B. Schatz. 1996. A Concept Space Approach to Addressing the Vocabulary Problem in Scientific Information Retrieval: An Experiment on the Worm Community System. *Journal of the American Society for Information Science*, 47(8).
- CONACyT (Consejo Nacional de Ciencia y Tecnología). 1998. Lista de Asignación 1997-1998 Fenómeno EL NIÑO.
- Ellis, C.A., S.J. Gibbs, and G.L. Rein, 1991. Groupware: Some Issues and Experiences. *Communications of the ACM*. 34(1), 39-58 p.
- Gosling, J., B. Joy and G. Steele. 1996. The Java Language Specification. First Edition. Addison-Wesley. Reading, Massachusetts. 825 pp
- Eltgroth, P.G., M.F. Wehner, A.A. Mirin, P.B. Duffy, K.G. Caldeira, J.H. Bolstad, H. Wang, C.M. Matarazzo and U.E. Creach. 1997. Comprehensive Climate System Modeling on Massively Parallel Computers. Proceedings of the Western Simulation Multiconference, Phoenix, AZ.
- Greenberg, S. 1990. Feasibility Study of a National High Speed Communications Networks for Research and Development: Future Applications. Research Report, Learning and Collaboration Group, Advanced Computing and Engineering Department, Alberta Research Council. 55 pp.
- Grudin, J., 1994. Computer-Supported Cooperative Work: History and Focus. *IEEE Computer*. 27(5). 19-26 p.

- IAI (The Inter-American Institute for Global Change Research). 1994. Report on the IAI Workshop on Enso and Interannual Climate Variability.  
<http://www.geo.nsf.gov/iai/iaiw8.htm>
- IAI (The Inter-American Institute for Global Change Research). 1997. 1996-97 Report Inter-American Institute for Global Change Research.  
[http://www.geo.nsf.gov/iai/annual\\_reports/iaia9697\\_sp.htm](http://www.geo.nsf.gov/iai/annual_reports/iaia9697_sp.htm)
- Isaacs, H., J. Tang and T. Morris. 1996. "Piazza: A Desktop Environment Supporting Impromptu and Planned Interactions". CSCW'96. Boston, MA. 315-324 p.
- Kim, W. 1990. Introduction to Object-Oriented Databases. First Edition. The MIT Press. Cambridge, Massachusetts. 234 pp.
- Kouzes, R. T., J.D. Myers and W.A. Wulf. 1996. Collaboratories: Doing Science on the Internet. IEEE Computer. 29(8). 40-46 p.
- Kraut, R. Y., J. Galegher. and C. Egido. 1988. Relationships and Tasks in Scientific Research Collaborations. Human Computer Interaction, 3(1), 31-58 p.
- Lederberg, J. 1989. Excitement and Fascination of Science. Third Edition. Palo Alto, California, Annual Reviews, Inc.
- Lewandowski, S. M. 1998. Frameworks for Component-Based Client/Server Computing. ACM Computing Surveys. 30(1). 3-27 p.
- Long-Term Ecological Research (LTER). 1998. <http://LTERnet.edu/>
- Malone, T. and K. Rockart. 1991 Computers, Networks and the Corporation. Scientific American. 265(3): 92-99 p.
- National Center for Biotechnology Information (NCBI). 1998. GenBank Overview  
<http://www.ncbi.nlm.nih.gov/Web/Genbank/index.html>
- National Center for Supercomputing Applications (NCSA). 1993. NCSA HDF Specifications and Developer's Guide. Champaign, Illinois. 74 pp
- National Center for Supercomputing Applications (NCSA). 1996. The HDF WWW Scientific Data Browser(CGI). <http://hdf.ncsa.uiuc.edu:4321/>
- National Center for Supercomputing Applications (NCSA). 1997. HDF User's Guide Version 4.1r1. Champaign, Illinois. 100 pp
- National Center for Supercomputing Applications (NCSA). 1988. Information, Support, and Software from the Hierarchical Data Format (HDF) Group of NCSA.  
<http://hdf.ncsa.uiuc.edu/>
- National Space Science Data Center (NSSDC). 1998. The Common Data Format.  
[http://nssdc.gsfc.nasa.gov/cdf/cdf\\_home.html](http://nssdc.gsfc.nasa.gov/cdf/cdf_home.html)
- National Oceanic and Atmospheric Administration (NOAA). 1998. National Geophysical Data Center. <http://www.ngdc.noaa.gov/>
- National Research Council (NRC). 1995. Preserving Scientific Data on Our Physical Universe: A New Strategy for Achieving the Nation's Scientific Information Resources. National Academy Press. First Edition. National Academy Press. Washington, D.C. 80 pp.

- National Research Council (NRC). 1997. Bits of Power: Issues in Global Access to Scientific Data. First Edition. National Academy Press. Washington, D.C. 250 pp.
- Nielsen, J. 1994. Guerrilla HCI: Using Discount Usability Engineering to Penetrate the Intimidation Barrier. En Bias, R. G. and D.J. Mayhew (Eds.), Cost-Justifying Usability. Academic Press, Boston, MA, 245-272 p.
- Orfali, R. and D. Harkey. 1998. Client/Server Programming with Java and CORBA. Second Edition. John Wiley & Sons. New York, NY. 1022 pp.
- Pressman, R. 1997. Software Engineering A Practitioners's Approach. Fourth Edition. McGraw-Hill. New York, NY. 852 pp.
- Roseman, M., and S. Greenberg. 1992. GoupKit: A Groupware Toolkit for Building Real-Time Conferencing Applications. ACM Conference on Computer Supported Cooperative Work (CSCW'92). Toronto, Nov. 1-4, ACM Press, 43-50 p.
- Soriano, T. 1996. Caracterización del Espacio de Diseño de Sistemas Colaborativos y su Aplicación en el Desarrollo de un Sistema Síncrono para la Revisión de Diagramas Técnicos. Tesis de Maestría. CICESE. Ensenada, Baja California.
- Stefik, M., G. Foster, D.G. Bobrow, K. Kahn, S. Lanning. and L. Suchman. 1987. Beyond the Chalkboard: Computer Support for Collaboration and Problem Solving in Meetings. Communications of the ACM. 30(1). 32-47 p.
- Sun Microsystems, Inc. (Sun). 1998. Java Remote Method Invocation Specification. <http://www.javasoft.com/products/jdk/1.2/docs/guide/rmi/spec/rmiTOC.doc.html>
- Tanenbaum, A. 1996. Computers Networks. Third Edition. Prentice Hall. Upper Saddle River, New Jersey. 814 pp
- University Corporation for Atmospheric Research (UCAR). 1998. Unidata netCDF. <http://www.unidata.ucar.edu/packages/netcdf>
- Ullman, J.D. 1988. Principles of Databases and Knowledge-Base Systems Vol. 1. Primera Edición. Computer Science Press. Rockville, Maryland. 631 pp.
- University of Connecticut (UConn). 1998. BIOTA: Biodiversity Database Manager <http://viceroy.eeb.uconn.edu/biota>
- Watson, J.D. 1990. The Human Genome Project: Past, Present, and Future. Science 248. 44-51 p.
- Westbrook, J.H. 1992. A History of Data Recording, Analysis, and Dissemination. Data for Discovery: Proceedings of the Twelfth International CODATA Conference, Columbus, Ohio, Jul. 15-19, 1990. 430-460 p.
- World Meteorological Organization (WMO). 1994. Guide to WMO Binary Code Forms. <http://www.wmo.ch/web/www/reports/Guide-binary.html>
- Zapata, M. 1997. Metodología de Diseño de Sistemas Colaborativos Síncronos y su Aplicación en el Desarrollo de un Sistema de Administración de Redes de Cómputo. Tesis de Maestría. CICESE. Ensenada, Baja California.