



CONSORZIO PER LA GESTIONE DEL CENTRO DI  
COORDINAMENTO DELLE ATTIVITÀ DI RICERCA INERENTI IL  
SISTEMA LAGUNARE DI VENEZIA

# Programma di Ricerca 2004-2006

## *Linea 4.2*

### Sistema informativo distribuito

**TITOLO DEL PROGETTO DI RICERCA:** Modellazione, analisi e visualizzazione di dati ambientali

**RESPONSABILE SCIENTIFICO:** Renzo Orsini

**COORDINATORE OPERATIVO:** Renzo Orsini

**ISTITUZIONI DEL GRUPPO DI RICERCA:**

Tipologia	Istituzione
Partecipante	Dipartimento di Informatica dell'Università Ca' Foscari di Venezia (DI-UNIVE)
Partecipante	Laboratoire d'InfoRmatique en Images et Systèmes d'information - Institut National des Sciences Appliquées de Lyon (LIRIS-INSA)

# 1 Descrizione della Ricerca

## 1.1 Descrizione in lingua italiana

I molti studi attualmente condotti sulla laguna di Venezia, anche nell'ambito delle ricerche finanziate dal CORILA, producono una quantità di informazioni scientifiche ambientali che si caratterizzano per la ricchezza, la vastità e l'eterogeneità. Affinché questo grande volume di informazioni sia pienamente utilizzabile allo scopo di approfondire le conoscenze sui processi che maggiormente contribuiscono alla contaminazione dell'ambiente lagunare, in modo da intraprendere opportune decisioni di controllo e di governo, sono indispensabili tre condizioni: 1) che fra dati di natura tanto diversa possano essere stabiliti il maggior numero di collegamenti e integrazioni possibili, 2) che vengano individuati strumenti di analisi molto raffinata, rivolta alla estrazione di vari tipi di conoscenze, da applicare non solo sulle singole serie di dati, ma sull'insieme più ampio possibile di informazioni collegate, e 3) che siano resi disponibili strumenti che facilitino l'accesso e la consultazione sia ai dati integrati che ai risultati delle varie analisi su di essi.

Questo progetto ha l'obiettivo di arricchire l'attuale sistema informativo del CORILA attraverso lo sviluppo di potenti strumenti per la modellizzazione integrata, l'analisi e l'accesso ai dati attualmente raccolti dai progetti di ricerca finanziati o disponibili da sorgenti di dati esterne, sia utilizzando i più recenti risultati della ricerca nelle aree dell'informatica relative a tali obiettivi, sia sviluppando nuova ricerca applicata al settore dell'elaborazione dei dati ambientali.

Il lavoro è suddiviso in tre *workpackage*, rivolti ai tre aspetti sopra menzionati.

Nel primo verrà affrontato il problema della modellazione dei dati ambientali e della loro integrazione a partire da sorgenti eterogenee. Il problema verrà affrontato da tre diversi punti di vista: a) definizione di meccanismi di strutturazione di dati efficaci per la memorizzazione in maniera integrata di dati di natura diversa, compresi dati di tipo continuo, estendendo gli attuali strumenti usati dal CORILA (come la base di dati RIVELA); b) individuazione di ontologie in ambito di scienze ambientali, con la produzione di strumenti per facilitarne l'uso ai ricercatori dell'area; c) predisposizione di strumenti per l'integrazione di dati ambientali provenienti da sorgenti eterogenee. I risultati di questa ricerca permetteranno di: a) sviluppare basi di dati integrate, da cui estrarre dati per analisi e costruzione di modelli che tengano in considerazione i collegamenti fra i dati; b) diffondere nell'area della ricerca ambientale utilizzata dal CORILA l'uso di ontologie di riferimento standard (i cosiddetti metadati), che permettano una migliore condivisione di informazioni sia non strutturate, come rapporti, relazioni, ecc., sia parzialmente o completamente strutturate, e una più efficace ricerca delle informazioni; c) avere a disposizione strumenti per lo sviluppo di modelli e software di analisi su dati memorizzati in sorgenti diverse ed eterogenee.

Nel secondo *workpackage* verranno studiate diverse tecniche innovative specifiche per l'analisi dei dati ambientali con lo scopo di fornire supporto ai processi decisionali. Una prima attività si focalizzerà su tecniche di *data mining* per l'estrazione di *pattern* e regole ricorrenti a partire da dati ambientali referenziati in spazio e/o tempo e su metodi per rappresentare la conoscenza estratta e per ragionare su questa. Si analizzeranno inoltre tecniche innovative per l'analisi di immagini provenienti da applicazioni per l'ambiente e l'architettura. In particolare verranno impiegate tecniche di riconoscimento di forme per arricchire le relative basi dati, e permettere nuovi tipi di ricerche soprattutto in ambito della conservazione dei beni culturali. Infine, un'altra attività riguarderà la modellazione, con l'approccio dei Sistemi Complessi, di vari componenti della realtà lagunare per valutarne le proprietà costituenti quali efficienza e punti deboli.

Il terzo *workpackage* è dedicato allo sviluppo di modelli e strumenti per l'analisi dei dati continui, dati cioè che modellano fenomeni continui e sono derivati da campi continui. Tali dati sono generalmente conosciuti solo in alcuni punti come dati campionati. Si ha quindi il problema di costruire interpolazioni spazio-temporali per conoscere il fenomeno in ogni punto, e di memorizzare non solo i punti campionati con le loro coordinate ed i loro attributi, ma anche i metodi d'interpolazione nel tempo e nello spazio. Saranno studiati tre aspetti: la rappresentazione dei dati continui, l'interrogazione di questi dati con integrazione automatica del metodo di interpolazione, e la loro visualizzazione.

## **1.2 Abstract in lingua inglese**

The research about the lagoon of Venice, including the studies financed by CORILA, produces a large amount of scientific environment-related information, characterized by richness, amplitude and heterogeneity. Such a huge volume of information must be made fully usable for deepening the knowledge on the processes that mainly contribute to the contamination of the lagoon environment, so as to undertake opportune decisions of control and government. Therefore, three conditions are necessary: 1) the largest number of connections and possible integrations between environmental data need to be established, 2) rich analysis tools must be designed or selected, targeted to the extraction of several types of knowledge, to be applied not only on single series of data, but also on the widest set of related information, and 3) tools must be made available that facilitate the access and the consultation of the integrated data and of the analyses results.

The goal of this project is to enrich the CORILA information system through the development of powerful tools for the integrated modeling, the analysis and the access to data collected from the currently financed research plans, or available from external sources, using the most recent results of computer science research and developing new applied research in the field of the environmental data processing.

The project is divided in three workpackages, oriented to the three aspects cited above.

The first workpackage will approach the problem of environmental data modeling and integration from heterogeneous data sources. The problem will be faced from three points of view: a) definition of efficient data structuring mechanisms for storing in an integrated way data of different types, including continuous data; in order to accomplish this task, the tools and systems currently used by CORILA (such as the RIVELA database) will be extended; b) development of ontologies for the environment science, and of tools for assisting the researchers in using them; c) development or selection of tools for the integration of environmental data from heterogeneous sources. The results of this research will lead to: a) develop integrated databases from which data can be extracted for analysis and construction of models exploiting the connections among them; b) disseminate in the area of environmental research approached by CORILA, standard ontologies (the so-called metadata), allowing researchers to better share unstructured information, such as reports, as well as semistructured and structured information, and to retrieve information efficiently; c) access instruments for developing models and analysis tools for data stored in several heterogeneous sources.

The second workpackage will study innovative techniques for the analysis of environmental data, aiming at supporting the decisional processes. A first activity will be focused on techniques for data mining, i.e., for the extraction of patterns and recurrent rules from data referenced in space and time, and on methods for representing the extracted knowledge and reasoning on it. Innovative techniques for image analysis in environmental and architectural applications will be studied. Shape recognition techniques will be used for enriching the databases with new information retrieval tools, mostly in the cultural heritage domain. Another activity will concern the modeling, with the complex systems approach, of several components of the lagoon scenario in order to estimate the constituent properties such as efficiency and weak aspects.

The third workpackage is dedicated to the development of models and tools for the analysis of continuous data, i.e., data modeling continuous phenomena and deriving from continuous fields. Such data are generally known only in some points as sampled data, and a continuous field must be reconstructed by time and space interpolation. In order to know the phenomenon in its entirety, the interpolation methods must be stored together with the sampled points. The workpackage will study three aspects: the representation of the continuous data, the querying of such data with automatic integration of the interpolation method, and their visualization.

## 1.3 Descrizione dettagliata della ricerca

### WP1

#### 1.3.1 Anno 2004

Verrà svolta un'analisi delle tecniche, degli standard e delle metodologie di modellazione e rappresentazione di dati di tipo ambientale, sia discreti che continui, anche in collegamento con il lavoro del workpackage 3. Si analizzeranno le eventuali iniziative di standardizzazione delle ontologie e dei *thesauri* nel settore dei dati ambientali. Verrà valutata la possibilità che la comunità scientifica coinvolta in progetti coordinati dal CORILA promuova o partecipi a sforzi congiunti sulla definizione di una ontologia di settore. Verrà fatta una classificazione delle principali fonti di dati ambientali sulla laguna di Venezia disponibili, per predisporre gli strumenti per una successiva integrazione.

#### 1.3.2 Anno 2005

Verrà definito e realizzato uno strumento basato su Web, se già non disponibile, per permettere ai ricercatori la creazione e la consultazione di una ontologia di settore. Verrà definita l'architettura e il linguaggio di specifica del sistema di integrazione di dati eterogenei, e verrà realizzato un prototipo.

#### 1.3.3 Anno 2006

Verrà definito uno schema di una base di dati compatibile con l'attuale base di dati RIVELA del CORILA ma che permetta l'integrazione di nuovi tipi di dati ambientali e di estensioni procedurali, come modelli e algoritmi di analisi. Sarà costruito lo schema della base di dati e il relativo modello fisico. Verrà stimolato l'uso dell'ontologia di riferimento nella comunità dei ricercatori ambientali sia attraverso opportune forme di comunicazione sia modificando gli strumenti di raccolta di dati scientifici usati dal CORILA per rendere vincolanti questo tipo di specifiche. Verrà consolidato il prototipo di integrazione dei dati eterogenei e verrà applicato alla basi di dati esterne disponibili.

### WP2

#### 1.3.1 Anno 2004

Verrà studiato lo stato dell'arte sugli algoritmi di *data mining* per l'estrazione di *pattern* da dati referenziati sia in spazio e sia in tempo, con particolare enfasi al caso *streaming* e distribuito. Nel contempo si analizzerà il dominio applicativo, ovvero il tipo di conoscenza estraibile dalle sorgenti di dati ambientali del CORILA. Verrà inoltre svolta un'indagine preliminare finalizzata all'individuazione di tecniche per la rappresentazione del dominio applicativo e della conoscenza estratta.

Verrà svolto uno studio sullo stato dell'arte relativa ai problemi di estrazione e classificazione di "image templates" in database di immagini, con rappresentazioni di tipo "landmark methods" e metodi di classificazione di tipo "machine learning".

Verrà studiata l'applicazione dell'approccio a Sistemi Complessi sulle varie tipologie di dati presenti nell'ambiente lagunare.

### 1.3.2 Anno 2005

Verrà completato lo studio dell'arte delle tecniche di *data mining* focalizzati sui dati CORILA su cui effettuare l'analisi.

Verranno sviluppati e analizzati algoritmi e applicazioni per il riconoscimento e l'estrazione di elementi architettonici di interesse da database di immagini in possesso del CORILA. Lo scopo sarà quello di classificare gli elementi estratti in classi predefinite di interesse per l'analisi e la conservazione dei beni architettonici, allo scopo di arricchire la relativa base di dati e permettere ricerche complesse.

Verrà studiato il modeling tramite Sistemi Complessi dei vari sistemi componenti la realtà lagunare, e costruiti degli strumenti software per la generazione di tali modelli.

### 1.3.3 Anno 2006

Verrà realizzato il prototipo di *data mining*, anche reso accessibile tramite Web, per l'analisi di dati ambientali. I dati che allo stato attuale risultano quantitativamente significativi per l'analisi sono quelli raccolti in ambito "Architettura e Beni culturali", eventualmente arricchiti dagli strumenti di riconoscimento di forme sviluppati in questo WP. Ci si riserva la possibilità di utilizzare anche, qualora si rendessero disponibili, dati ambientali di tipo continuo, referenziati in spazio e tempo, raccolti nel sistema di datawarehouse sviluppato nell'ambito del WP3, oppure in altri sistemi resisi disponibili.

Verrà sviluppato del software per analizzare gli aspetti strutturali della realtà lagunare vista come Sistema Complesso, in particolare l'efficienza dei sistemi, ed i loro punti deboli.

## **WP3**

### 1.3.1 Anno 2004

Si definirà un modello per la rappresentazione, elaborazione e query di dati continui campionati. Utilizzando approcci di tipo data warehouse, ben noti nell'ambito dei sistemi discreti, dove i dati sono organizzati in strutture gerarchiche, si introdurranno meccanismi di interpolazione sulle dimensioni spazio-temporali per dati continui. Si proporrà quindi un'estensione del modello multidimensionale ai dati continui. Si studieranno gli ambiti di applicabilità compatibili con i dati disponibili presso il CORILA.

Si studierà inoltre il problema dell'interazione visuale con dati georeferenziati proponendo un modello user-oriented.

### **1.3.2 Anno 2005**

Si studierà il problema di interrogare un data warehouse contenente dati continui estendendo le tecnologie OLAP (On-Line Analytical Processing) e Spatial OLAP (SOLAP) ai dati interpolati. Si estenderanno gli operatori classici e si definiranno nuovi operatori per facilitare all'utente la navigazione nei dati continui e per trovare rappresentazioni adeguate dei dati continui. Si affronterà il problema dell'interrogazione e della visualizzazione dei dati continui realizzando alcuni prototipi.

Si proseguirà nell'analisi del problema dell'interazione visuale e si definirà un protocollo di validazione di usabilità che sarà verificato sui prototipi sviluppati.

### **1.3.3 Anno 2006**

L'attività sarà soprattutto sperimentale. Utilizzando e perfezionando i prototipi realizzati di un sistema per la gestione di dati continui in un data warehouse, di un sistema di interrogazione visuali e di strumenti per la visualizzazione, si analizzeranno i dati forniti da CORILA su un caso di studio specifico.

Si completerà l'analisi di usabilità sui prototipi di interazione visuale e si proporranno le modifiche necessarie per migliorarli.

## **1.4 Risultati attesi**

### **WP1**

#### **1.4.1 Anno 2004**

La prima fase del WP1 ci darà la conoscenza sia delle strutture delle basi di dati ambientali, che dei metadati usati per classificare conoscenze in questo settore, per quanto è pubblicamente disponibile. La raccolta delle informazioni sulle sorgenti dei dati ambientali accessibili sulla laguna di Venezia costituirà un primo repertorio di informazioni da mettere a disposizione della comunità dei ricercatori. Entrambe queste attività forniranno una prima indicazione per rispondere al quesito 1 del bando.

#### **1.4.2 Anno 2005**

Il prototipo di interrogazione di sorgenti dati eterogenee servirà come base per identificare, descrivere, e iniziare a risolvere le problematiche principali relative al quesito 1 del bando. Lo sviluppo degli strumenti per un'ontologia di settore permetterà di sviluppare un'ontologia di riferimento comune all'integrazione delle diverse basi di dati.

#### **1.4.3 Anno 2006**

Il nuovo schema di basi di dati definito renderà disponibile un modello di una base di dati integrata che ci permetterà di rispondere ad una parte del quesito 1 del bando. Il completamento

del prototipo di integrazione permetterà l'accesso alle basi di dati disponibili fornendo una capacità di visualizzazione integrata, completando quindi la risposta al quesito 1 del bando.

## WP2

### 1.4.1 Anno 2004

La prima fase del WP2 ci darà una prima conoscenza per quanto riguarda gli strumenti di analisi per l'estrazione di conoscenza da dati spazio-temporali, per la loro rappresentazione, e per la loro validazione modellistica. Verranno inoltre approfonditi i metodi di analisi di pattern presenti in immagini di tipo architettonico, con particolare attenzione ai metodi basati sui "landmark". Verrà infine acquisita la necessaria conoscenza di dominio, ovvero le caratteristiche e la struttura dei dati in possesso del CORILA in relazione alle possibili tecniche di analisi utilizzabili. Si farà inoltre un'analisi del sistema laguna (tramite i dati disponibili e in fase di acquisizione del CORILA) dal punto di vista dei Sistemi Complessi, valutando le potenzialità proprie dei vari sottosistemi e proponendo un ranking opportuno. Queste attività iniziali all'interno del WP ci permetteranno di fornire una prima indicazione per rispondere ai quesiti 2, 3 e alla prima parte del quesito 5 del bando.

### 1.4.2 Anno 2005

Mediante lo sviluppo di strumenti per la segmentazione di immagini, l'estrazione di elementi architettonici e la loro classificazione, sarà possibile fornire risposte alla prima parte del quesito 5 del bando.

Si produrranno, in accordo con i risultati sviluppati nel 2004, modellizzazioni di sottosistemi lagunari, usando la teoria dei Sistemi Complessi, contribuendo alla risposta al quesito 3 del bando.

### 1.4.3 Anno 2006

L'attività del WP2 in questo ultimo anno porterà alla realizzazione di un prototipo per l'estrazione e l'analisi di conoscenza, che integra a livello software le operazioni per l'estrazione di *pattern* frequenti e regole e permette quindi di ragionare su dati spazio-temporali. Questo, oltre allo sviluppo e alla sperimentazione di algoritmi per l'estrazione, la classificazione e l'indicizzazione di elementi architettonici presenti in un database di immagini, ci consentirà di completare nel corso dell'anno le risposte ai quesiti 2 e 5 del bando. In oltre questi patterns potrebbero essere utilizzati per il controllo di qualità delle misure (sensori bloccati, ecc.).

Si svolgerà un'analisi delle proprietà rilevanti dei Sistemi Complessi lagunari, studiando proprietà essenziali quali efficienza e centralità delle risorse. Si produrrà inoltre uno studio dei punti deboli/critici di tali Sistemi Complessi, contribuendo alla risposta al quesito 3 del bando.

## WP3

### 1.4.1 Anno 2004

La prima fase del WP3 produrrà un modello per la rappresentazione dei campi continui che aggiungerà ulteriori contributi a quanto prodotto dal WP1 per rispondere al quesito 1 del bando. Lo studio della modellazione dei campi continui attraverso un approccio data warehouse fornirà indicazioni preliminari anche per rispondere al quesito 3 del bando.

Lo studio dei requisiti degli utenti per quanto riguarda l'interazione visuale con i dati georeferenziati permetterà di comprendere la specificità dell'ambiente CORILA e di progettare opportunamente i sistemi di interazione con le applicazioni (quesito 4 del bando).

### 1.4.2 Anno 2005

Si imposterà un sistema Spatial OLAP per l'interpolazione e l'analisi dei dati incompleti secondo un approccio data warehouse, contribuendo a fornire risposta al quesito 3 del bando.

Si otterranno anche prototipi per l'interazione con gli algoritmi e i programmi di interpolazione di campi continui e una metodologia per la verifica della loro usabilità, rispondendo così ai quesiti 3 e 4 del bando.

### 1.4.3 Anno 2006

Si completeranno i programmi di gestione e ricostruzione dei campi continui mediante interpolazione in modo da completare, insieme ai risultati del WP1, la risposta al quesito 1 del bando.

L'implementazione di prototipi di sistemi di query visuali e visualizzazione dei risultati per i campi continui consentirà di accedere in modo efficace ai dati gestiti dal CORILA. Per valutarne l'efficacia si attueranno delle verifiche di usabilità con gli utenti, che consentiranno di completare le risposte al quesito 4 e agli aspetti di visualizzazione dei dati del quesito 1 del bando.