ArchaeoGRID
a Grid for Archaeology

P.G.Pelfer

Univ. Florence and INFN, Florence

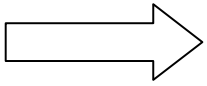
EGEE User Forum, CERN, March 1-3, 2006

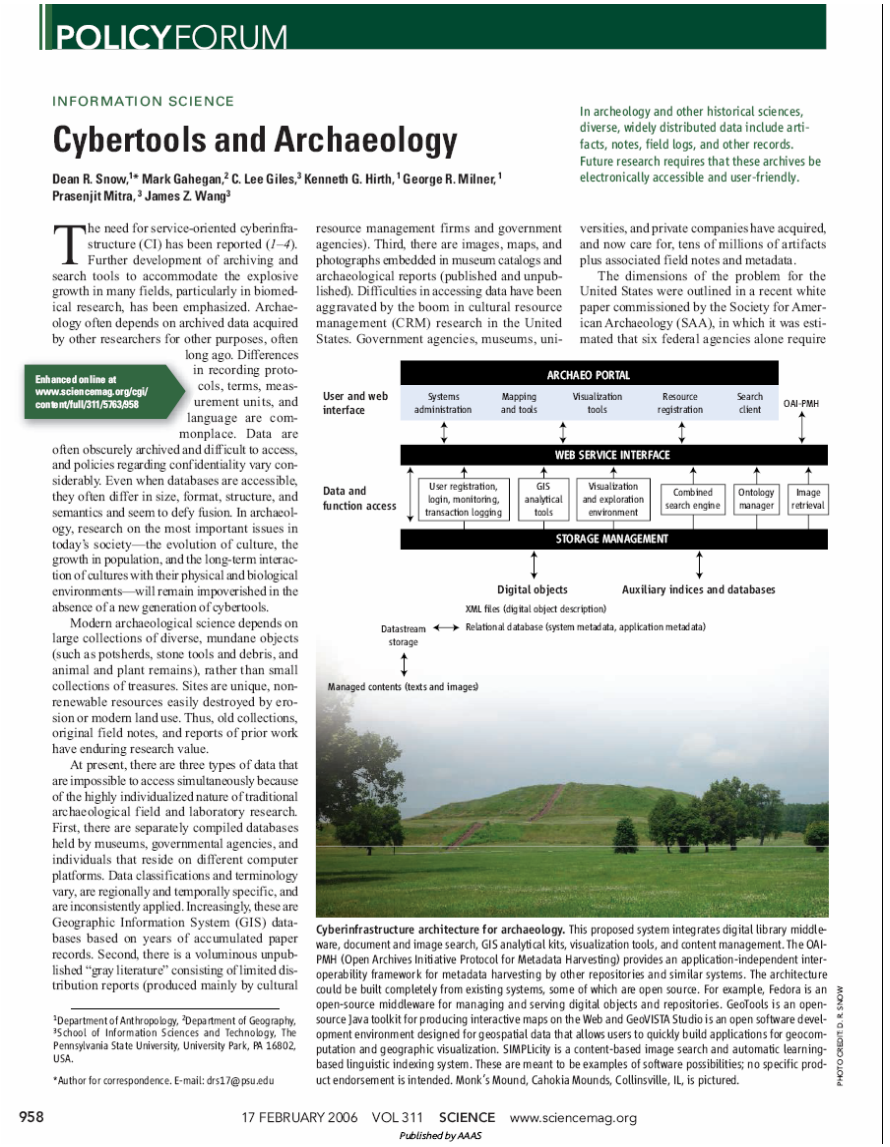
www.eu-egee.org

Information Society and Media

Short History of Idea

- **2003**
  - G.Pelfer, P.G.Pelfer *"From Web to GRID, a new perspective for Archaeology"* poster, NSS-MIC IEEE Conference, 2003, Portland, USA
- **2004**
  - G.Pelfer, P.G.Pelfer et al. *"ArchaeoGRID, a Grid for Archaeology"*, NSS-MIC IEEE Conference, 2004, Rome
- **2005**
  - P.G.Pelfer, EGAAP Committee, Pisa, 2005
- **2006**
  - 17 February 2006





POLICY FORUM

INFORMATION SCIENCE

Cybertools and Archaeology

Dean R. Snow,[1]* Mark Gahegan,[2] C. Lee Giles,[3] Kenneth G. Hirth,[1] George R. Milner,[1] Prasenjit Mitra,[3] James Z. Wang[3]

In archeology and other historical sciences, diverse, widely distributed data include artifacts, notes, field logs, and other records. Future research requires that these archives be electronically accessible and user-friendly.

The need for service-oriented cyberinfrastructure (CI) has been reported (*1*–*4*). Further development of archiving and search tools to accommodate the explosive growth in many fields, particularly in biomedical research, has been emphasized. Archaeology often depends on archived data acquired by other researchers for other purposes, often long ago. Differences in recording protocols, terms, measurement units, and language are commonplace. Data are often obscurely archived and difficult to access, and policies regarding confidentiality vary considerably. Even when databases are accessible, they often differ in size, format, structure, and semantics and seem to defy fusion. In archaeology, research on the most important issues in today's society—the evolution of culture, the growth in population, and the long-term interaction of cultures with their physical and biological environments—will remain impoverished in the absence of a new generation of cybertools.

Enhanced online at www.sciencemag.org/cgi/content/full/311/5763/958

Modern archaeological science depends on large collections of diverse, mundane objects (such as potsherds, stone tools and debris, and animal and plant remains), rather than small collections of treasures. Sites are unique, nonrenewable resources easily destroyed by erosion or modern land use. Thus, old collections, original field notes, and reports of prior work have enduring research value.

At present, there are three types of data that are impossible to access simultaneously because of the highly individualized nature of traditional archaeological field and laboratory research. First, there are separately compiled databases held by museums, governmental agencies, and individuals that reside on different computer platforms. Data classifications and terminology vary, are regionally and temporally specific, and are inconsistently applied. Increasingly, these are Geographic Information System (GIS) databases based on years of accumulated paper records. Second, there is a voluminous unpublished "gray literature" consisting of limited distribution reports (produced mainly by cultural resource management firms and government agencies). Third, there are images, maps, and photographs embedded in museum catalogs and archaeological reports (published and unpublished). Difficulties in accessing data have been aggravated by the boom in cultural resource management (CRM) research in the United States. Government agencies, museums, universities, and private companies have acquired, and now care for, tens of millions of artifacts plus associated field notes and metadata.

The dimensions of the problem for the United States were outlined in a recent white paper commissioned by the Society for American Archaeology (SAA), in which it was estimated that six federal agencies alone require

Cyberinfrastructure architecture for archaeology. This proposed system integrates digital library middleware, document and image search, GIS analytical kits, visualization tools, and content management. The OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting) provides an application-independent interoperability framework for metadata harvesting by other repositories and similar systems. The architecture could be built completely from existing systems, some of which are open source. For example, Fedora is an open-source middleware for managing and serving digital objects and repositories. GeoTools is an open-source Java toolkit for producing interactive maps on the Web and GeoVISTA Studio is an open software development environment designed for geospatial data that allows users to quickly build applications for geocomputation and geographic visualization. SIMPLicity is a content-based image search and automatic learning-based linguistic indexing system. These are meant to be examples of software possibilities; no specific product endorsement is intended. Monk's Mound, Cahokia Mounds, Collinsville, IL, is pictured.

PHOTO CREDIT: D. R. SNOW

[1]Department of Anthropology, [2]Department of Geography, [3]School of Information Sciences and Technology, The Pennsylvania State University, University Park, PA 16802, USA.

*Author for correspondence. E-mail: drs17@psu.edu

EGAAP Recommendations

- ……
- **2) Archeogrid**
- **The proposal to include applications from human sciences in the applications EGEE portfolio is very attractive. EGAAP is therefore minded to recommend full approval for Archeogrid at its next meeting March 1-3, if the following actions are taken:**
- **The ARCHEOGRID community should approach the DILIGENT collaboration since many questions raised in Archeogrid are common to the DILIGENT program of work.**
- **The ARCHEOGRID community should deploy on GILDA their paleoclimate application.**
- **The present EGEE infrastructure is not suited for the moment for the other proposed application using remote sensors.**
- **A discussion should take place with the earth sciences community to evaluate potential synergies between the climate applications and the paleoclimate ones.**
- **The ARCHEOGRID collaboration in terms of material and human resources and expertise should be better described.**
- ……

Archaeological Data

- Modern archaeological science depends on large collections of diverse, mundane objects - such as potsherds, stone tools and debris, and animal and plant remains - rather than small collections of treasures.
- Sites are unique, nonrenewable resources easily destroyed by erosion or modern land use.
- Old collections, original field notes, and reports of prior work have enduring research value

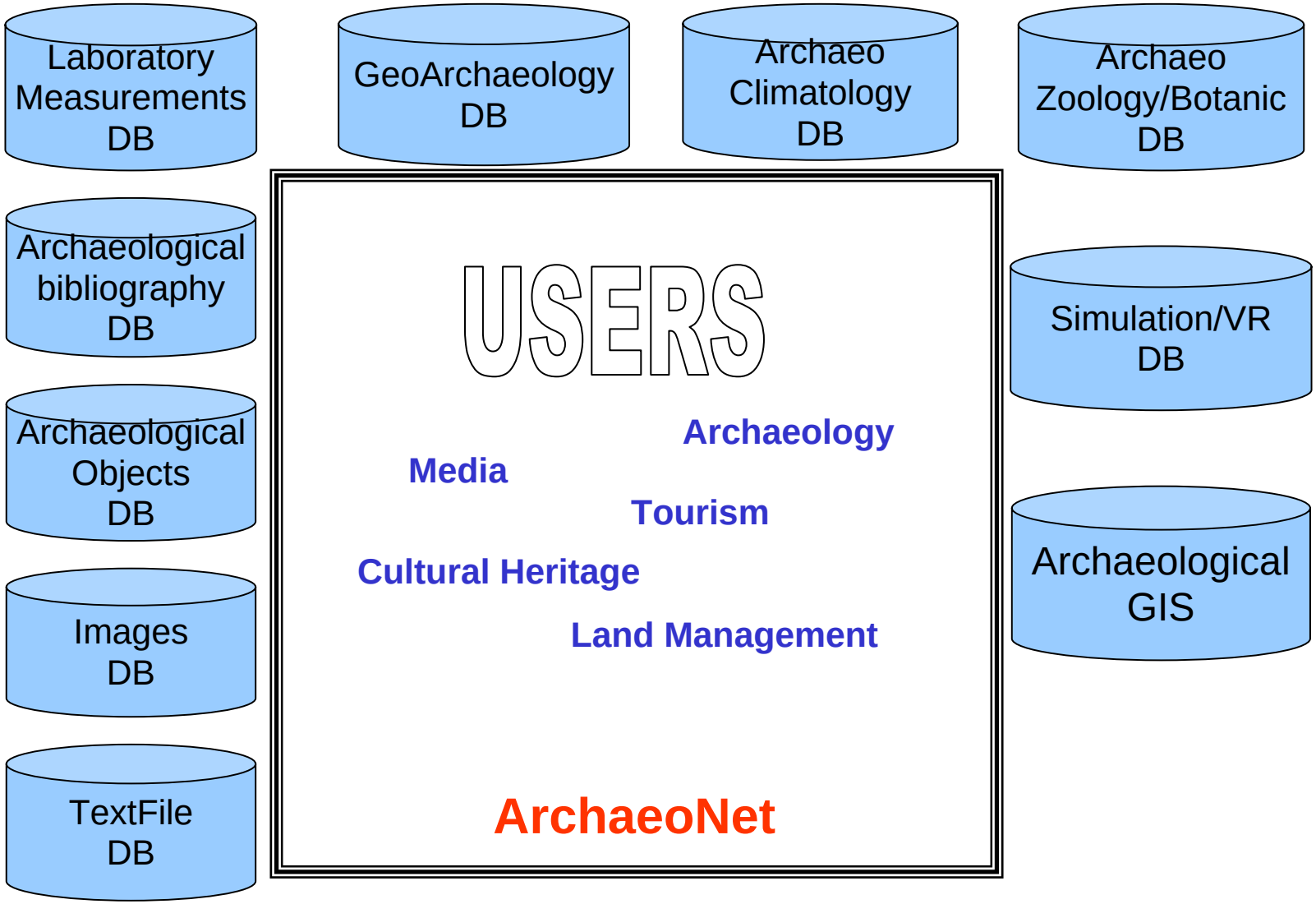
Archaeological Data

- Huge amount of data, with not easy simultaneous access because of the highly individualized nature of traditional archaeological field and laboratory research, can be classified in three different types:
  - separately compiled databases held by museums, governmental agencies, private companies and individuals that reside on different computer platforms. Data classifications and terminology vary, regionally and temporally, and are inconsistently applied. Many Archaeological GIS databases based on years of accumulated paper records.
  - voluminous unpublished literature consisting of limited distribution reports.
  - images, maps, and photographs embedded in museum catalogues and archaeological reports, published and unpublished.
- **Archaeological data are geospatial data**

Archaeology and Internet

- Emergence of **e-Archaeology Global Knowledge Community** for exploitation of advanced Computational methods to generate, curate and analyse research data and to develop and explore models and simulation
- **Why the Archaeogrid ?**
  - for research in archaeology on the most important issues into the evolution of ancient societies: demographic problems; access to resources; long-term interaction of cultures with physical and biological environments; context extended from the site to large regions
  - for the increasing complexity of activities involving archaeology - research, education, archaeological heritage preservation and access, media, land management, tourism
  - **for an easy access to widespread data and analysis tools. In Italy about 2000 archaeological parks and museums.**

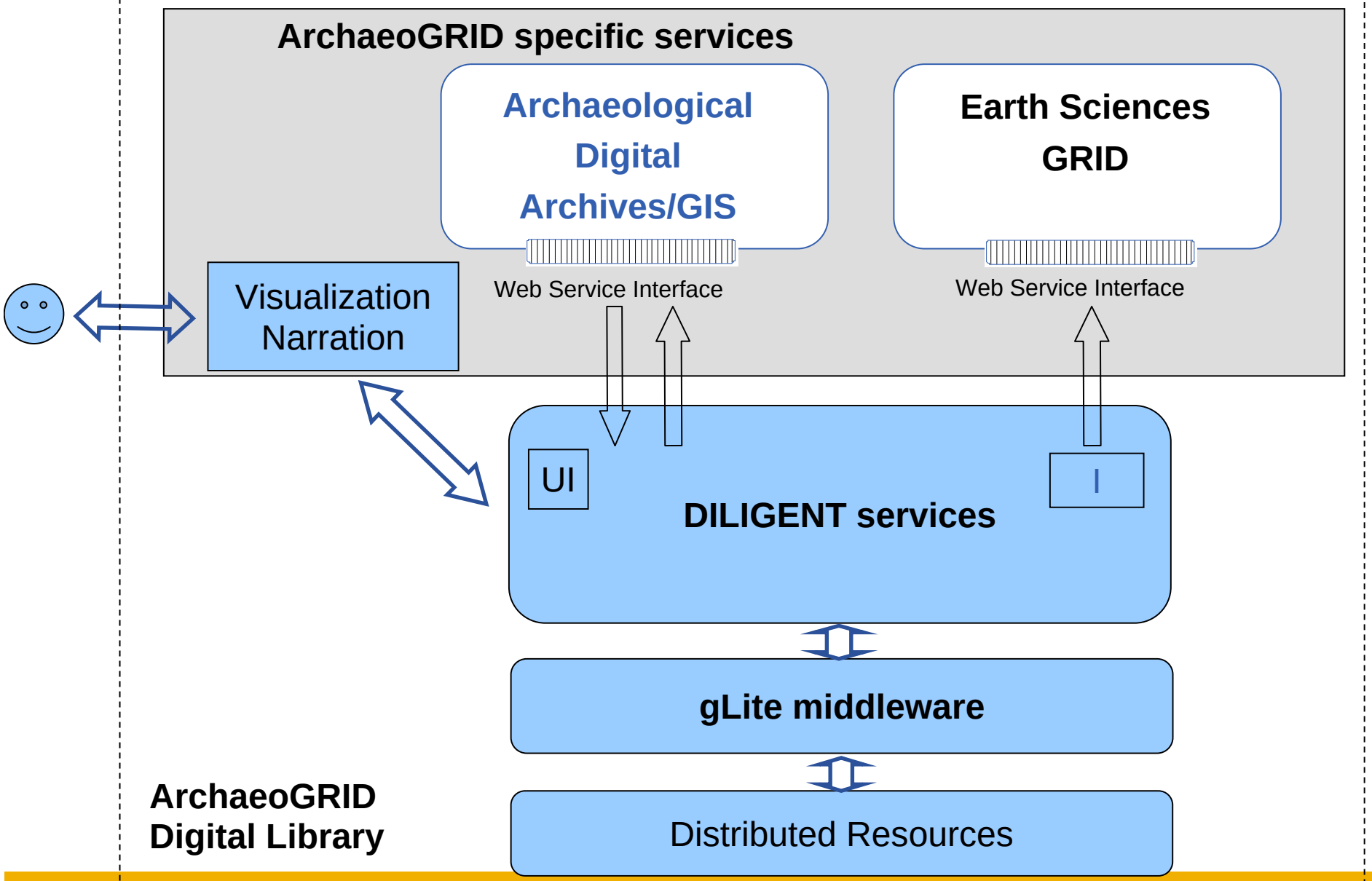
Archaeological Information and Users



Archaeological Activities

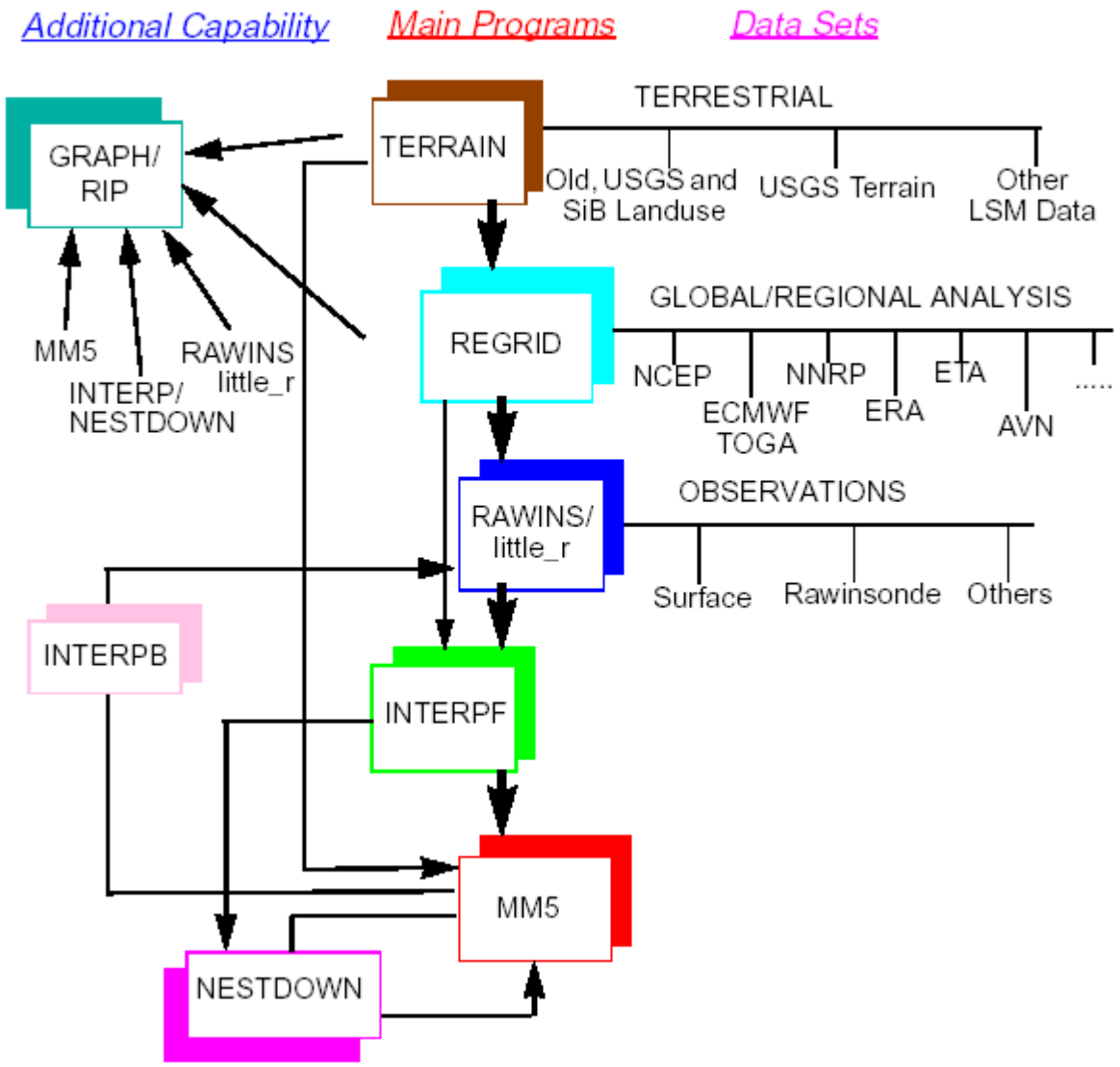
- **Archaeological Activities**
  - **Data acquisition**
    - Field Survey and Landscape Archaeology
    - Excavations
    - Laboratory Analysis
    - Reconstruction of Process and Sites – “Experimental Archaeology”
  - **Data storage and analysis**
  - **Archaeological “document” production**
    - Visualization
    - Narration
- **Archaeological historical documents are composed by maps, images, draws, tables, etc. linked by text. Textual parts, “narration”, connect all the elements of the documents in a coherent vision based on a “theoretical model”.**

ArchaeoGRID Digital Library



GILDA Paleoclimate Application

MM5 Modeling System Flow Chart

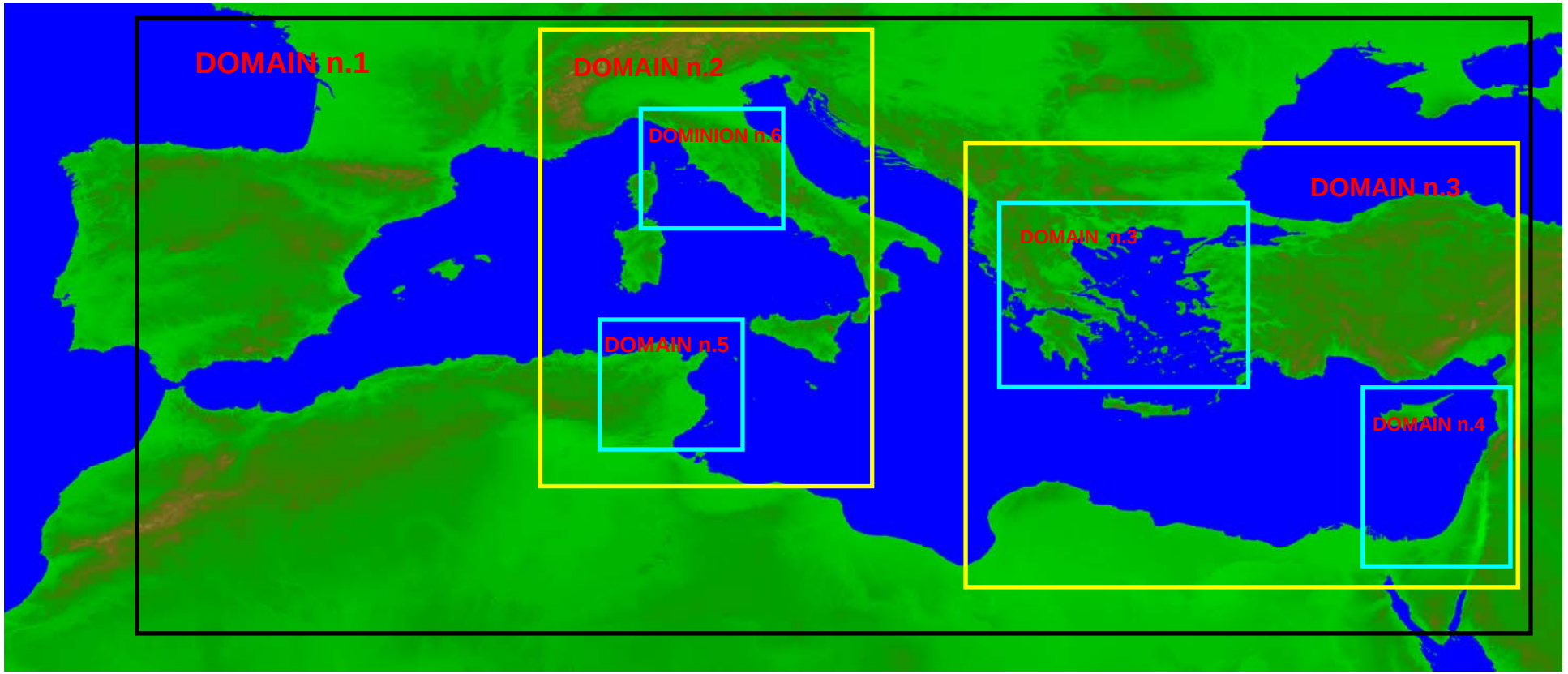


MM5 Mesoscale Model on GILDA

- **Software and Data on GILDA Infrastructure**
- **MM5, Mesoscale Model**
  - already used for archaeological paleoclimate simulations
  - source code and data available
  - drawbacks
    - MPI
    - proprietary Linux fortran compilers (pgf90, ifort)
- **ES Paleoclimatology vs Archaeoclimatology**
  - simulation in Archaeoclimatology can use model and data from ES Paleoclimatology if spatial and temporal resolution of Model and data are enough precise for Archaeological Applications
  - data from ES Paleoclimatology could be integrated with data from archaeological survey and simulation (DEM, land cover, soils, etc.)

MM5 Domains

The origin of City during the X-VIII centuries B.C.
in Mediterranean area and the MM5 paleoclimatic simulations
- DTM mosaic from GTOPO30 data -



Group Composition and Contacts

- **Pier Giovanni PELFER, University of Florence and INFN, Italy, Application Supervisor, pelfer@fi.infn.it**
- **Giuliano PELFER, Center for Study of Complex Dynamics, University of Florence, Italy, VO Manager, pelfer@tin.it**
- **Roberto CECCHINI, INFN, Florence, Local Coordinator, Italy, Roberto.Cecchini@fi.infn.it**
- **Antonio POLITI, Institute for Complexity Study (ISC), CNR, Florence, Italy, Local Coordinator,**
- **Mercedes FARJAS, Polytechnic University of Madrid (UPM), Madrid, Spain, Local coordinator, m.farjas@upm.es**
- **Oscar BLAZQUEZ, CEHTEX Company, Madrid, Spain, Industrial Partner, oscarblazquez@cehtex.es**
- **Apostolos SARRIS, Institute for Mediterranean Studies, FORTH, Crete, Greece, Local Coordinator, asaris@ims.forth.gr**

New contacts will start, related to specific archaeological applications.

ArchaeoGRID Expertise

- **Italy: Grid technology, GIS, Spatial Analysis, Multi-Agent Based Modeling applied to Archaeology, Landscape Archaeology**
- **Spain: Geodesy ,Topography, Cartography and GIS for Archaeology and Cultural Resource Management, Paleoenvironment**
- **Greece: Paleoclimate and Paleoenvironment Simulations, Remote Sensing for Archaeology, Geophysical Archaeological Prospection, GIS, Landscape Archaeology**

ArchaeoGRID Synergies

- GILDA Group in Catania full supported and support the ArchaeoGRID activity. GILDA Infrastructure is the place where continue to run archaeological application like paleoclimate, paleoenvironment and multi-agent based simulations.
- DILIGENT Group develop a technology needed for the ArchaeoGRID. The ArchaeoGRID Digital Library approach appear as the best approach. ArchaeoGRID schema have been discussed with DILIGENT ISTI CNR Pisa Group.
- ES Paleoclimate: e-mail contacts with IPSL (Monique Petididier) and with DKRZ (Kerstin Ronnenberger, Uwe Mikolajewicz, Daniela Jacob) . It seems possible to use Paleoclimate Model running on EGEE Infrastructure, but it needs a deeper discussion about the archaeoclimatology approach.