UNIVERSITÀ DI PISA

DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

Dottorato di Ricerca in Ingegneria dell'Informazione

Corso di Dottorato

"Using e-Infrastructures for Biodiversity Conservation"

Dr. Gianpaolo Coro

Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo" (ISTI) - CNR- Italy

Short Abstract: An e-Infrastructure is a distributed network of service nodes, residing on multiple sites and managed by one or more organizations. e-Infrastructures allow scientists residing at distant places to collaborate. They offer a multiplicity of facilities as-a-service, supporting data sharing and usage at different levels of abstraction, e.g. data transfer, data harmonization, data processing workflows etc. e-Infrastructures are gaining an important place in the field of biodiversity conservation. Their computational capabilities help scientists to reuse models, obtain results in shorter time and share these results with other colleagues. They are also used to access several and heterogeneous biodiversity catalogues.

In this course, the D4Science e-Infrastructure will be used to conduct experiments in the field of biodiversity conservation. D4Science hosts models and contributions by several international organizations involved in the biodiversity conservation field. The course will give students an overview of the models, the practices and the methods that large international organizations like FAO and UNESCO apply by means of D4Science. At the same time, the course will introduce students to the basic concepts under e-Infrastructures, Virtual Research Environments, data sharing and experiments reproducibility.

Course Contents in brief:

- e-Infrastructures and for Virtual Research Environments
- Practice with the D4Science e-Infrastructure
- Geospatial data visualization and representation
- Statistical models for species distribution modelling
- Accessing to large heterogeneous biodiversity catalogs
- Signal processing of biodiversity-related observations
- Machine Learning applied to species observation records
- Lexical search in large taxonomic trees
- Cloud computing applied to biodiversity analyses

Total # of hours: 20 (5 credits)

References:

[1] Candela L., Castelli D., Coro G., Pagano P., Sinibaldi F. Species distribution modeling in the cloud. In: Concurrency and Computation-Practice & Experience, Geoffrey C. Fox, David W. Walker, Ed. Wiley, DOI: 10.1002/cpe.3030

[2] G. Coro, P. Pagano, A. Ellenbroek, ”Combining Simulated Expert Knowledge with Neural Networks to Produce Ecological Niche Models for Latimeria chalumnae”, Ecological Modelling, Ed. Elsevier, DOI 10.1016/j.ecolmodel.2013.08.005.

[3] Coro G., Pagano P., Ellenbroek A. Comparing Heterogeneous Distribution Maps for Marine Species. GIScience & Remote Sensing, Ed Taylor & Francis, DOI 10.1080/15481603.2014.959391.

[4] Candela L., Castelli D., Coro G., De Faveri F., Italiano A., Lelii L., Mangiacrapa F., Marioli V., Pagano P. Integrating Species Occurrence Databases to Facilitate Data Analysis. Ecological Informatics Journal, Elsevier, DOI 10.1016/j.ecoinf.2014.07.006.

[5] Coro G., Gioia A., Pagano P., Candela L. A service for statistical analysis of marine data in a distributed e-infrastructure. In: Bollettino di Geofisica Teorica e Applicata: an International Journal of Earth Sciences, vol. 54 (Suppl.) pp. 68 - 70. Supplement: IMDIS 2013 - International Conference on Marine Data and Information Systems, 23-25 September, Lucca (Italy). OGS - Istituto Nazionale di Oceanografia e di Geofisica Sperimentale, 2013.

[6] R. Froese, J. Thorson, R. B. Reyes Jr. A Bayesian approach for estimating length-weight relationships in fishes. Journal of Applied Ichthyology. Volume 30, Issue 1, pages 78–85, 2013

[7] Candela, L., Castelli, D., Pagano, P., 2009. D4Science: an e-Infrastructure for Supporting Virtual Research Environments. In: Agosti, M., Esposito, F., Thanos, C. (Eds.), Post-proceedings of the 5th Italian Research Conference on Digital Libraries - IRCDL 2009. DELOS: an Association for Digital Libraries, pp. 166–169.

CV of the Teacher

Gianpaolo Coro is a Physicist with a PhD in Computer Science. His research focuses on Artificial Intelligence and Data Mining. He has been working on Machine Learning and Signal Processing with applications to Computational Biology, Brain Computer Interfaces, Language Technologies and Cognitive Sciences. The aim of his research is the study and experimentation of models and methodologies to process biological data and to apply the results to fields in Ecological Modelling, Vessel Monitoring Systems and Ecological Niche Modelling. His approach relies on distributed e-Infrastructures and uses parallel and distributed computing via Grid and Cloud based technologies.

Room and Schedule

Room: *Aula Riunioni del Dipartimento di Ingegneria dell’Informazione, Via G. Caruso 16, Pisa – Ground Floor*

Schedule: May 2015

Day 1. Introduction to e-Infrastructures, Virtual Research Environments and Large Biodiversity Catalogs– 9.00 – 13.00

Day 2. Geospatial data descriptions, catalogs and visualization – 9.00 - 13.00

Day 3. Trends analysis of species observation records and environmental data– 9.00 - 13.00

Day 4. Data Processing: operations on large species datasets and taxonomic trees – 9.00 – 13.00

Day 5. Data Processing: species distribution modelling using machine learning techniques and Cloud computing– 9.00 – 13.00