

## Editorial

# Earth and Space Science Informatics: informatics in oceanography

G. M. R. Manzella<sup>1</sup>, S. Iona<sup>2</sup>, and S. Nativi<sup>3</sup>

<sup>1</sup>ENEA UTMAR, La Spezia, Italy

<sup>2</sup>HODC HCMR Athens, Greece

<sup>3</sup>CNR IMAA, Prato, Italy

A decade ago, information was discrete and disconnected, assembling data from many different sources was inherently time consuming, and was made for specific studies, such as evaluation of trends in ecosystems, climatology productions, etc. Most of the data were not directly accessible, but were sent by mail on CD-ROMs or floppy discs.

The world is changed and the collection of digital data is growing at an exponential rate, making the access to data more difficult than before. During the last decade, with the advent of Internet-based system architectures a fundamental shift has occurred.

The digitization of all kinds of data, the implementation of computer networks and the availability of broad band have posed the basis for the connection of a vast amount of data and information. Internet, Web, distributed information systems and computing infrastructures have been changing the working way in science, engineering, business, and education. Today, advances in all kind of application domains is often related with the possibilities to access heterogeneous information and the related tools to understand and use them.

Different scientific and technological communities are today working together to develop federated information systems by applying interoperability solutions and recognising that:

- Data in the Earth sciences is being used by a far broader community than ever before.
- Information harvesting by selecting the appropriate, exchange metadata, the evolving interoperability standards, and the multi-disciplinary scenarios raises important and difficult challenges.
- An effort done by most of information technology projects is the provision of information describing data

collection, processing and use –i.e. metadata. Digital libraries and simpler Wiki sites are among the solution adopted.

- Metadata are critical for recording the provenance of data, including processing steps and quality control certification (i.e. *who, what, where, when, which, why, how*).
- A basic international goal is the definition of a standard set of metadata and a mechanism for exchanging the acquired information.

Metadata was a word rarely mentioned in the past, but they are now critical for recording the provenance of datasets. Once published, important datasets should remain available and be used after the end of the projects that gathered them. Thus, to understand datasets a later user needs to discover and access the metadata (description) on:

1. *how* the acquisition instruments were designed and built;
2. *what* was acquired;
3. *when, where* and *how* the datasets were gathered and pre-processed ;
4. a careful description of the processing steps (*how*);
5. *who* acquired and processed the datasets;
6. *why* the datasets were acquired and processed

The subject of metadata definition in this context is not only related to the metadata encoding , but also to the semantic aspects of this metadata, as it is of prime importance that the definition of a value at metadata level has the same meaning for all interoperable data repositories, which is not obvious at our starting point, as each data repository system



Correspondence to: S. Nativi  
(stefano.nativi@cnr.it)

has achieved its own development lifecycle in its “cultural” project environment.

The development and adoption of clearer dataset descriptions, common data structures, documented and common data processing, and standard discovery and access protocols will improve data exchange and, as a consequence, provide more benefits to science and society. There are still many barriers to the provision of dataset to intermediate and end users, some of them being related to information technology (e.g. encoding, data model, semantics).

Due to the diverse information technologies and methodologies, interoperability has been defined as one of the priorities in order to provide the end-users with an easier access to different kinds of products. Interoperability is not a new concept in the information science. The idea has always been to ensure that data and information can flow, as efficiently as possible, between people and systems, how easily different kinds of software systems work together. Hence, data managed from multiple heterogeneous data sources need to be combined requiring complex processing such as content mediation and aggregation. This is not a simple task, because of the different legacies that characterize diverse data management and processing systems.

The evolution of the Information Systems raises many issues of trust, concerning: security, confidentiality, ownership, assured provenance, authenticity, and information quality.

The new alliance between Earth science and information technology has been beneficial to the society, transforming the data-centric concepts and architectures into Service-Oriented Architectures (SOA). This change has also involved the concepts of data storage, giving a major importance to data preservation for future use. The importance of data preservation is never underlined enough. In fact, the following principles must be carefully considered:

- Use and re-use of data can provide new services for the civil society.
- Access to data can generate new studies, including collection-based research to generate new science.
- Retention of unique observational data which is impossible to re-create.
- Retention of expensively generated data which is cheaper to maintain than to re-generate.
- Data are necessary to assess compliance with legal requirements.
- Data are necessary to validate published research results and for use in teaching.

The papers published in this volume are providing some important guidelines for an interoperable infrastructure (e-Infrastructure) in terms of: data and metadata models, network services, application schemas, catalogues, semantics, security, data policy, etc. They represent significant efforts that are undergoing in the Oceanography discipline, and more generally in the Earth Science System Informatics (ESSI) domain. Presently, they are among the most challenging research fields for an informed society.

*Acknowledgements.* This publication was financially supported by the European Commission DG Research, Environment Directorate.