Toward open science at the European scale: geospatial semantic array programming for integrated environmental modelling

Daniele de Rigo and Paolo Corti and Giovanni Caudullo and Daniel McInerney and Margherita Di Leo and Jesús San-Miguel-Ayanz

(1) European Commission, Joint Research Centre, Institute for Environment and Sustainability, (2) Politecnico di Milano, Dipartimento di Elettronica e Informazione, (3) United Nations World Food Programme

2013

Online at https://mpra.ub.uni-muenchen.de/45958/
MPRA Paper No. 45958, posted 8 April 2013 12:56 UTC
Interfacing science and policy raises challenging issues when large spatial-scale (regional, continental, global) environmental problems need transdisciplinary integration within a context of modelling complexity and multiple sources of uncertainty [1]. This is characteristic of science-based support for environmental policy at European scale [1], and key aspects have also long been investigated by European Commission transnational research [2–5].

\[
\text{Geospatial data } \ X = \{X_1 \cdots X_n\} \quad \text{(raw, derived information)}
\]

\[
\begin{align*}
\text{Remote sensing} \\
\text{Scattered time series and field observations} \\
\text{Statistics over territorial administrative units} \\
\text{Raster/vectorial derived data}
\end{align*}
\]

\( (a.1) \quad (a.2) \quad (a.3) \quad (a.4) \)

\[
\text{Parameters of the needed data-transformations } \ \theta = \{\theta_1 \cdots \theta_m\}
\]

\( (a.5) \)
Wide-scale transdisciplinary modelling for environment

Approaches (either of computational science or of policy-making) suitable at a given domain-specific scale may not be appropriate for wide-scale transdisciplinary modelling for environment (WSTM) and corresponding policy-making [6–10]. In WSTM, the characteristic heterogeneity of available spatial information (a) and complexity of the required data-transformation modelling (D-TM) appeal for a paradigm shift in how computational science supports such peculiarly extensive integration processes. In particular, emerging wide-scale integration requirements of typical currently available domain-specific modelling strategies may include increased robustness and scalability along with enhanced transparency and reproducibility [11–15]. This challenging shift toward open data [16] and reproducible research [11] (open science) is also strongly suggested by the potential – sometimes neglected – huge impact of cascading effects of errors [1,14,17–19] within the impressively growing interconnection among domain-specific computational models and frameworks.

From a computational science perspective, transdisciplinary approaches to integrated natural resources modelling and management (INRMM) [20] can exploit advanced geospatial modelling techniques with an awesome battery of free scientific software [21,22] for generating new information and knowledge from the plethora of composite data [23–26].

From the perspective of the science-policy interface, INRMM should be able to provide citizens and policy-makers with a clear, accurate understanding of the implications of the technical apparatus on collective environmental decision-making [1]. Complexity of course should not be intended as an excuse for obscurity [27–29].

Array Programming [30] (b)

\[
\begin{align*}
\text{array based D-TM} & \quad f(X, \theta) \\
\text{data-dependent parameters (sub D-TM)} & \quad \theta(X) \\
\text{array based semantics} & \quad \text{=}
\end{align*}
\]

GNU Octave [31,32] (MATLAB language)
- concise support for large complex valued multidimensional D-TM, sparse matrices, nested mixed arrays, higher order functions (b.1)

GNU R [33] (R language)
- wide libraries of statistical tests, data analysis, classification, clustering (b.2)

GNU Bash [34]
- commandline robust and scalable tools
  for concise text and file based D-TM,
  scripting (sed, grep, awk, GNU Core Utilities, ... ) (b.3)

Mastrave [35,36] (MATLAB language, GNU Bash, )
- Semantic Array Programming,
  support for array based functional programming (b.4)

Python [37] (Numpy [38], Scipy [39])
- Array-oriented (e.g. geo-layers) Javascript libraries
  concise interface with geo-tools (c) and data (a) (b.5)

...
Geospatial Semantic Array Programming

Concise array-based mathematical formulation and implementation (with array programming tools, see (b)) have proved helpful in supporting and mitigating the complexity of WSTM [40–47] when complemented with generalized modularization and terse array-oriented semantic constraints. This defines the paradigm of Semantic Array Programming (SemAP) [35,36] where semantic transparency also implies free software use (although black-boxes [12] – e.g. legacy code – might easily be semantically interfaced).

A new approach for WSTM has emerged by formalizing unorganized best practices and experience-driven informal patterns. The approach introduces a lightweight (non-intrusive) integration of SemAP and geospatial tools (c) – called Geospatial Semantic Array Programming (GeoSemAP). GeoSemAP (d) exploits the joint semantics provided by SemAP and geospatial tools to split a complex D-TM into logical blocks which are easier to check by means of mathematical array-based and geospatial constraints. Those constraints take the form of precondition, invariant and postcondition semantic checks. This way, even complex WSTM may be described as the composition of simpler GeoSemAP blocks, each of them structured as (d).

\[ \text{Geospatial tools} = \begin{cases} \text{Systems for supporting geographic resources analysis (e.g. scriptable GIS such as GRASS GIS [48–50], ... )} & (c.1) \\ \text{Geospatial data abstraction library (GDAL [51])} & (c.2) \\ \text{Geospatial web support (e.g. with OGC WPS [52]: pyWPS, OpenLayers [53], ... )} & (c.3) \\ \text{Geospatial database support (e.g. scriptable data queries with PostGIS [54] by using (b.3) [55], ... )} & (c.4) \\ \text{...} & \end{cases} \]
GeoSemAP allows intermediate data and information layers to be more easily and formally semantically described so as to increase fault-tolerance [17], transparency and reproducibility of WSTM_e. This might also help to better communicate part of the policy-relevant knowledge, often difficult to transfer from technical WSTM_e to the science-policy interface [1,15].

References


