Geophysical Research Abstracts
Vol. 15, EGU2013-13245-9, 2013
EGU General Assembly 2013

Toward Open Science at the European Scale: Geospatial Semantic Array Programming for Integrated Environmental Modelling

Daniele de Rigo (1,2), Paolo Corti (1,3), Giovanni Caudullo (1), Daniel McInerney (1), Margherita Di Leo (1), and Jesús San-Miguel-Ayanz (1)

(1) European Commission, Joint Research Centre, Institute for Environment and Sustainability, Via E. Fermi 2749, I-21027 Ispra (VA), Italy, (2) Politecnico di Milano, Dipartimento di Elettronica e Informazione, Via Ponzio 34/5, I-20133 Milano, Italy, (3) United Nations World Food Programme, Via C.G.Viola 68 Parco dei Medici, I-00148 Rome, Italy

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].

$$\textbf{(a)}\qquad \begin{array}{c}\text{Geospatial data}\\ X=\{X_1\cdots X_n\}\\ \text{(raw, derived information)}\end{array} = \begin{cases} \begin{array}{l}\text{Remote sensing}\\ \qquad\text{different spatial, spectral,}\\ \qquad\text{radiometric, temporal resolution}\end{array} & \text{(a.1)}\\[1ex] \begin{array}{l}\text{Scattered time series and field observations}\\ \qquad\text{e.g. irregular spatial density of sampling}\end{array} & \text{(a.2)}\\[1ex] \begin{array}{l}\text{Statistics over territorial administrative units}\\ \qquad\text{coarse spatial aggregation over irregular}\\ \qquad\text{polygons, e.g. NUTS, ISO 3166 - 2, ...}\end{array} & \text{(a.3)}\\[1ex] \begin{array}{l}\text{Raster/vectorial derived data}\\ \qquad\text{e.g. polygons describing focal}\\ \qquad\text{habitat patterns, regular grids of}\\ \qquad\text{categorical/numerical variables}\end{array} & \text{(a.4)}\\ \qquad\text{...} & \end{cases}$$

$$\text{Parameters of the needed data-transformations}\qquad \theta=\{\theta_1\cdots\theta_m\}\qquad\qquad \text{(a.5)}$$

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 (WSTMe) and corresponding policy-making [6–10]. In WSTMe, 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].

(b)

Array Programming [30]

array based D-TM $f(X,\theta)$

data-dependent parameters (sub D-TM) $\theta(X)$

array based semantics

=

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 WSTMe [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 WSTMe 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 WSTMe may be described as the composition of simpler GeoSemAP blocks, each of them structured as **(d)**.

(c)

Geospatial tools

geospatial D-TM, geospatial semantics

=

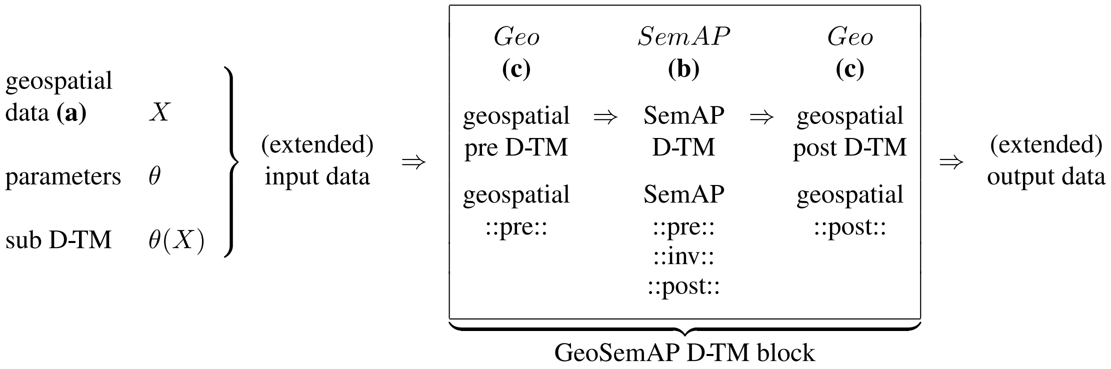
Systems for supporting geographic resources analysis (e.g. scriptable GIS such as GRASS GIS [48–50], ...) (c.1)

Geospatial data abstraction library (GDAL [51]) (c.2)

Geospatial web support (e.g. with OGC WPS [52]: pyWPS, OpenLayers [53], ...) (c.3)

Geospatial database support (e.g. scriptable data queries with PostGIS [54] by using (b.3) [55], ...) (c.4)

...



(d)

where

data	Data definition is extended to include proper geospatial data **(a)**, static parameters and sub D-TM – when used as dynamic (e.g. data-dependent) parameters
sub D-TM	Callback (function handle) to e.g. empirical equations, regression families, metrics/distance functions, ...
::pre::	Semantic pre-conditions
::inv::	Semantic invariants
::post::	Semantic post-conditions

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 WSTMe. This might also help to better communicate part of the policy-relevant knowledge, often difficult to transfer from technical WSTMe to the science-policy interface [1,15].

References

[1] de Rigo, D., 2013. Behind the horizon of reproducible integrated environmental modelling at European scale: ethics and practice of scientific knowledge freedom. F1000 Research. To appear as discussion paper.

[2] Funtowicz, S. O., Ravetz, J. R., 1994. Uncertainty, complexity and post-normal science. Environmental Toxicology and Chemistry 13 (12), 1881-1885. http://dx.doi.org/10.1002/etc.5620131203

[3] Funtowicz, S. O., Ravetz, J. R., 1994. The worth of a songbird: ecological economics as a post-normal science. Ecological Economics 10 (3), 197–207. http://dx.doi.org/10.1016/0921-8009(94)90108-2

[4] Funtowicz, S. O., Ravetz, J. R., 2003. Funtowicz, S., Ravetz, J. (2003). Post-normal science. International Society for Ecological Economics, Internet Encyclopaedia of Ecological Economics

[5] Ravetz, J., 2004. The post-normal science of precaution. Futures 36 (3), 347-357. http://dx.doi.org/10.1016/S0016-3287(03)00160-5

[6] van der Sluijs, J. P., 2012. Uncertainty and dissent in climate risk assessment: A Post-Normal perspective. Nature and Culture 7 (2), 174-195. http://dx.doi.org/10.3167/nc.2012.070204

[7] Ulieru, M., Doursat, R., 2011. Emergent engineering: a radical paradigm shift. International Journal of Autonomous and Adaptive Communications Systems 4 (1), 39-60. http://dx.doi.org/10.1504/IJAACS.2011.037748

[8] Turner, M. G., Dale, V. H., Gardner, R. H., Dec. 1989. Predicting across scales: Theory development and testing. Landscape Ecology 3 (3), 245-252. http://dx.doi.org/10.1007/BF00131542

[9] Zhang, X., Drake, N. A., Wainwright, J., 2004. Scaling issues in environmental modelling. In: Wainwright, J., Mulligan, M. (Eds.), Environmental modelling : finding simplicity in complexity. Wiley. ISBN: 9780471496182

[10] Bankes, S. C., 2002. Tools and techniques for developing policies for complex and uncertain systems. Proceedings of the National Academy of Sciences of the United States of America 99 (Suppl 3), 7263-7266. http://dx.doi.org/10.1073/pnas.092081399

[11] Peng, R. D., 2011. Reproducible research in computational science. Science 334 (6060), 1226-1227. http://dx.doi.org/10.1126/science.1213847

[12] Morin, A., Urban, J., Adams, P. D., Foster, I., Sali, A., Baker, D., Sliz, P., 2012. Shining light into black boxes. Science 336 (6078), 159-160. http://dx.doi.org/10.1126/science.1218263

[13] Nature, 2011. Devil in the details. Nature 470 (7334), 305-306. http://dx.doi.org/10.1038/470305b

[14] Stodden, V., 2012. Reproducible research: Tools and strategies for scientific computing. Computing in Science and Engineering 14, 11-12. http://dx.doi.org/10.1109/MCSE.2012.82

[15] de Rigo, D., Corti, P., Caudullo, G., McInerney, D., Di Leo, M., San-Miguel-Ayanz, J., (exp. 2013). Supporting Environmental Modelling and Science-Policy Interface at European Scale with Geospatial Semantic Array Programming. In prep.

[16] Molloy, J. C., 2011. The open knowledge foundation: Open data means better science. PLoS Biology 9 (12), e1001195+. http://dx.doi.org/10.1371/journal.pbio.1001195

[17] de Rigo, D., 2013. Software Uncertainty in Integrated Environmental Modelling: the role of Semantics and Open Science. Geophysical Research Abstracts 15, EGU General Assembly 2013.

[18] Cerf, V. G., 2012. Where is the science in computer science? Commun. ACM 55 (10), 5. http://dx.doi.org/10.1145/2347736.2347737

[19] Wilson, G., 2006. Where's the real bottleneck in scientific computing? American Scientist 94 (1), 5+. http://dx.doi.org/10.1511/2006.1.5

[20] de Rigo, D. 2012. Integrated Natural Resources Modelling and Management: minimal redefinition of a known challenge for environmental modelling. Excerpt from the Call for a shared research agenda toward scientific knowledge freedom, Maieutike Research Initiative. http://www.citeulike.org/groupfunc/15400/home

[21] Stallman, R. M., 2005. Free community science and the free development of science. PLoS Med 2 (2), e47+. http://dx.doi.org/10.1371/journal.pmed.0020047

[22] Stallman, R. M., 2009. Viewpoint: Why "open source" misses the point of free software. Communications of the ACM 52 (6), 31-33. http://dx.doi.org/10.1145/1516046.1516058 (free access version: http://www.gnu.org/philosophy/open-source-misses-the-point.html)

[23] Rodriguez Aseretto, D., Di Leo, M., de Rigo, D., Corti, P., McInerney, D., Camia, A., San Miguel-Ayanz, J., 2013. Free and Open Source Software underpinning the European Forest Data Centre. Geophysical Research Abstracts 15, EGU General Assembly 2013.

[24] Giovando, C., Whitmore, C., Camia, A., San-Miguel-Ayanz, J., 2010. Enhancing the European Forest Fire Information System (EFFIS) with open source software. In: FOSS4G 2010. http://2010.foss4g.org/presentations_show.php?id=3693

[25] Corti, P., San-Miguel-Ayanz, J., Camia, A., McInerney, D., Boca, R., Di Leo, M., 2012. Fire news management in the context of the European Forest Fire Information System (EFFIS). In: proceedings of "Quinta conferenza italiana sul software geografico e sui dati geografici liberi" (GFOSS DAY 2012). http://files.figshare.com/229492/Fire_news_management_in_the_context_of_EFFIS.pdf

[26] McInerney, D., Bastin, L., Diaz, L., Figueiredo, C., Barredo, J. I., San-Miguel-Ayanz, J., 2012. Developing a forest data portal to support Multi-Scale decision making. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 5 (6), 1-8. http://dx.doi.org/10.1109/JSTARS.2012.2194136

[27] Morin, A., Urban, J., Adams, P. D., Foster, I., Sali, A., Baker, D., Sliz, P., (2012). Shining light into black boxes. Science 336 (6078), 159-160. http://dx.doi.org/10.1126/science.1218263

[28] Stodden, V., 2011. Trust your science? Open your data and code. Amstat News July 2011, 21-22. http://www.stanford.edu/~vcs/papers/TrustYourScience-STODDEN.pdf

[29] van der Sluijs, J., 2005. Uncertainty as a monster in the science-policy interface: four coping strategies. Water Science & Technology 52 (6), 87-92. http://www.iwaponline.com/wst/05206/wst052060087.htm

[30] Iverson, K. E., 1980. Notation as a tool of thought. Communications of the ACM 23 (8), 444-465. http://awards.acm.org/images/awards/140/articles/9147499.pdf

[31] Eaton, J. W., Bateman, D., Hauberg, S., 2008. GNU Octave: a high-level interactive language for numerical computations. Network Theory. ISBN: 9780954612061

[32] Eaton, J. W., 2012. GNU octave and reproducible research. Journal of Process Control 22 (8), 1433-1438. http://dx.doi.org/10.1016/j.jprocont.2012.04.006

[33] R Development Core Team, 2011. The R reference manual. Network Theory Ltd. Vol. 1, ISBN: 978-1-906966-09-6. Vol. 2, ISBN: 978-1-906966-10-2. Vol. 3, ISBN: 978-1-906966-11-9. Vol. 4, ISBN: 978-1-906966-12-6.

[34] Ramey, C., Fox, B., 2006. Bash reference manual : reference documentation for Bash edition 2.5b, for Bash version 2.05b. Network Theory Limited. ISBN: 978-0-9541617-7-4.

[35] de Rigo, D., 2012. Semantic array programming for environmental modelling: Application of the mastrave library. In: Seppelt, R., Voinov, A. A., Lange, S., Bankamp, D. (Eds.), International Environmental Modelling and Software Society (iEMSs) 2012 International Congress on Environmental Modelling and Software. Managing Resources of a Limited Planet: Pathways and Visions under Uncertainty, Sixth Biennial Meeting. pp. 1167-1176. http://www.iemss.org/iemss2012/proceedings/D3_1_0715_deRigo.pdf

[36] de Rigo, D., 2012. Semantic Array Programming with Mastrave - Introduction to Semantic Computational Modelling. http://mastrave.org/doc/MTV-1.012-1.htm

[37] Van Rossum, G., Drake, F.J., 2011. Python Language Ref. Manual, Network Theory Ltd. ISBN: 0954161785. http://www.network-theory.co.uk/docs/pylang/

[38] The Scipy community, 2012. NumPy Reference Guide. SciPy.org. http://docs.scipy.org/doc/numpy/reference/

[39] The Scipy community, 2012. SciPy Reference Guide. SciPy.org. http://docs.scipy.org/doc/scipy/reference/

[40] de Rigo, D., Castelletti, A., Rizzoli, A. E., Soncini-Sessa, R., Weber, E., Jul. 2005. A selective improvement technique for fastening neuro-dynamic programming in water resources network management. In: Zítek, P. (Ed.), Proceedings of the 16th IFAC World Congress. Vol. 16. International Federation of Automatic Control (IFAC), pp. 7-12. http://dx.doi.org/10.3182/20050703-6-CZ-1902.02172

[41] de Rigo, D., Bosco, C., 2011. Architecture of a Pan-European Framework for Integrated Soil Water Erosion Assessment. Vol. 359 of IFIP Advances in Information and Communication Technology. Springer Boston, Berlin, Heidelberg, Ch. 34, pp. 310-318. http://dx.doi.org/10.1007/978-3-642-22285-6_34

[42] San-Miguel-Ayanz, J., Schulte, E., Schmuck, G., Camia, A., Strobl, P., Liberta, G., Giovando, C., Boca, R., Sedano, F., Kempeneers, P., McInerney, D., Withmore, C., de Oliveira, S. S., Rodrigues, M., Durrant, T., Corti, P., Oehler, F., Vilar, L., Amatulli, G., Mar. 2012. Comprehensive monitoring of wildfires in Europe: The European Forest Fire Information System (EFFIS). In: Tiefenbacher, J. (Ed.), Approaches to Managing Disaster - Assessing Hazards, Emergencies and Disaster Impacts. InTech, Ch. 5. http://dx.doi.org/10.5772/28441

[43] de Rigo, D., Caudullo, G., San-Miguel-Ayanz, J., Stancanelli, G., 2012. Mapping European forest tree species distribution to support pest risk assessment. In: Baker, R., Koch, F., Kriticos, D., Rafoss, T., Venette, R., van der Werf, W. (Eds.), Advancing risk assessment models for invasive alien species in the food chain: contending with climate change, economics and uncertainty. Bioforsk FOKUS 7. OECD Co-operative Research Programme on Biological Resource Management for Sustainable Agricultural Systems; Bioforsk - Norwegian Institute for Agricultural and Environmental Research. http://www.pestrisk.org/2012/BioforskFOKUS7-10_IPRMW-VI.pdf

[44] Estreguil, C., Caudullo, G., de Rigo, D., Whitmore, C., San-Miguel-Ayanz, J., 2012. Reporting on European forest fragmentation: Standardized indices and web map services. IEEE Earthzine. http://www.earthzine.org/2012/07/05/reporting-on-european-forest-fragmentation-standardized-indices-and-web-map-services/

[45] Estreguil, C., de Rigo, D. and Caudullo, G. (exp. 2013). Towards an integrated and reproducible characterisation of habitat pattern. Submitted to Environmental Modelling & Software

[46] Amatulli, G., Camia, A., San-Miguel-Ayanz, J., 2009. Projecting future burnt area in the EU-mediterranean countries under IPCC SRES A2/B2 climate change scenarios (JRC55149), 33-38

[47] de Rigo, D., Caudullo, G., Amatulli, G., Strobl, P., San-Miguel-Ayanz, J. (exp. 2013). Modelling tree species distribution in Europe with constrained spatial multi-frequency analysis. In prep.

[48] GRASS Development Team, 2012. Geographic Resources Analysis Support System (GRASS) Software. Open Source Geospatial Foundation. http://grass.osgeo.org http://www.spatial-ecology.net/dokuwiki/doku.php?id=wiki:firemod

[49] Neteler, M., Bowman, M. H., Landa, M., Metz, M., 2012. GRASS GIS: A multi-purpose open source GIS. Environmental Modelling & Software 31, 124-130. http://dx.doi.org/10.1016/j.envsoft.2011.11.014

[50] Neteler, M., Mitasova, H., 2008. Open source GIS a GRASS GIS approach. ISBN: 978-0-387-35767-6

[51] Warmerdam, F., 2008. The geospatial data abstraction library. In: Hall, G. B., Leahy, M. G. (Eds.), Open Source Approaches in Spatial Data Handling. Vol. 2 of Advances in Geographic Information Science. Springer Berlin Heidelberg, pp. 87-104. http://dx.doi.org/10.1007/978-3-540-74831-1_5

[52] Open Geospatial Consortium, 2007. OpenGIS Web Processing Service version 1.0.0. No. OGC 05-007r7 in OpenGIS Standard. Open Geospatial Consortium (OGC). http://portal.opengeospatial.org/files/?artifact_id=24151

[53] Hazzard, E., 2011. Openlayers 2.10 beginner's guide. Packt Publishing. ISBN: 1849514127

[54] Obe, R., Hsu, L., 2011. PostGIS in Action. Manning Publications. http://dl.acm.org/citation.cfm?id=2018871

[55] Sutton, T., 2009. Clipping data from postgis. linfiniti.com Open Source Geospatial Solutions. http://linfiniti.com/2009/09/clipping-data-from-postgis/