

# Chapter 1

## Ecological Informatics: An Introduction

Friedrich Recknagel and William K. Michener

### 1.1 Introduction

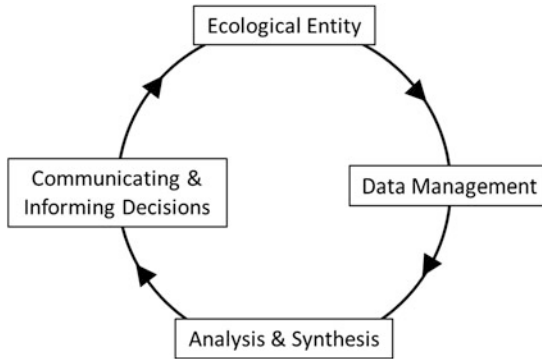
Ecological Informatics is an emerging discipline that takes into account the data-intensive nature of ecology, the valuable information content of ecological data, and the need to communicate results and inform decisions, including those related to research, conservation and resource management (Recknagel 2017). At its core, ecological informatics combines developments in information technology and ecological theory with applications that facilitate ecological research and the dissemination of results to scientists and the public. Its conceptual framework links ecological entities (genomes, organisms, populations, communities, ecosystems, landscapes) with data management, analysis and synthesis, and communicating and informing decisions by following the course of a loop (Fig. 1.1).

*Ecological Entities* range from genomes, individual organisms, populations, communities, ecosystems to landscapes and the biosphere, and are highly complex and distinctly evolving. Figure 1.2 illustrates the evolving nature of ecosystems in view of the fact that physical-chemical boundaries such as topology, temperature, pH, and substrate determine their community of organisms. Progressing shifts of physical-chemical boundaries under the influence of environmental and climate changes at seasonal and inter-annual scales restructure communities of organisms, and ecosystems adjust in due course. Over time, evolving ecosystems also alter the nature of landscapes. Ecologists are challenged by the evolving nature and

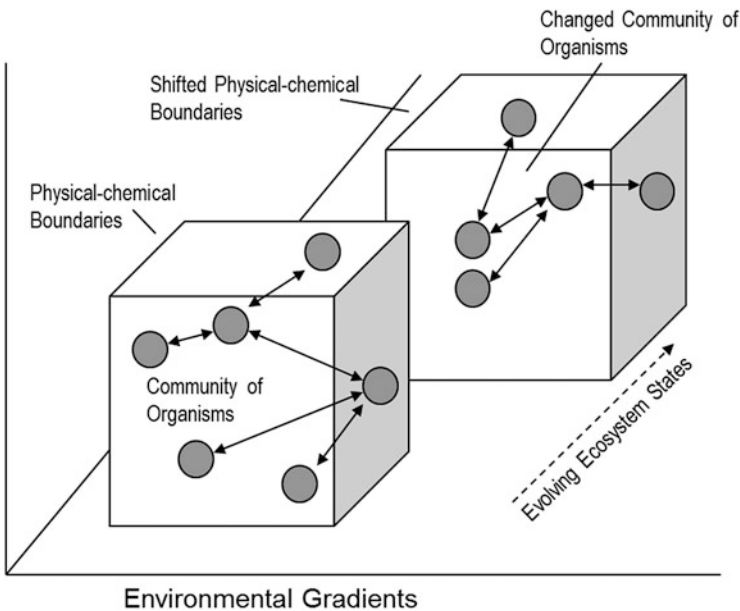
---

F. Recknagel (✉)  
University of Adelaide, Adelaide, SA, Australia  
e-mail: [friedrich.recknagel@adelaide.edu.au](mailto:friedrich.recknagel@adelaide.edu.au)

W.K. Michener  
The University of New Mexico, Albuquerque, NM, USA  
e-mail: [william.michener@gmail.com](mailto:william.michener@gmail.com)



**Fig. 1.1** Conceptual framework of ecological informatics



**Fig. 1.2** Evolving nature of ecosystems

data-intensive nature of ecology, and require suitable concepts and tools to deal appropriately with these challenges.

*Data Management* must meet requirements of many diverse sources of information, and be suitable to a wide range of spatial and temporal scales. Sources of information include paleo-ecological, eco-genomic, habitat, community and climate data. Spatial scales of ecological data range from habitat-specific to global, and time scales range from real-time to centuries-long.

*Analysis and Synthesis* utilise archived and real-time information for inventorying ecological entities, assessing sustainability of habitats and biodiversity, and

hind- and forecasting of ecological entities. Multivariate statistics are commonly applied for data analysis. Data synthesis typically applies inferential and process-based modelling techniques, and utilises remote sensing and GIS-based tools. Bayesian inference extends the predictive capacity of inferential and process-based models by quantifying model uncertainties and estimating forecasting risks.

*Communicating and Informing Decisions* supported by data analysis and synthesis is relevant for generating hypotheses for subsequent research steps as well as for identifying viable management options. While inferential models help inform short-term decisions, process-based models are more appropriate for long-term forecasts and decision-making.

## 1.2 Data Management

Ecological data management is a process that starts at the conceptualization of the project and concludes after the data have been archived and the results have informed future research as well as resource management, conservation, and other types of decision-making. Data management may be conceptualized in terms of a data life cycle (Fig. 1.3) whereby: (1) projects are conceived and data collection and analyses are planned; (2) data are collected and organized, usually into data tables (e.g., spreadsheets) or databases; (3) data are quality assured using accepted quality assurance/quality control (QA/QC) techniques; (4) data are documented through the creation of metadata that describe all aspects of the data and research; (5) data are preserved in a data repository or archive so that they may be reused and shared; (6) data are discovered or made discoverable so that they may be used in synthesis efforts or to reproduce results of a study; (7) data are integrated

**Fig. 1.3** The life cycle of data. Note the steps need not be sequential nor does research necessarily involve all steps; e.g., some synthesis efforts may involve no new data collection, thereby proceeding from data discovery through integration with other data, to analysis and visualization



with other data in order to answer specific questions such as examining the influence of climate extremes on pollination ecology; and (8) data are explored, analysed and visualized, leading to new understanding that can then be communicated to other scientists and the public.

The seven chapters in Part II discuss concepts, practices and tools that are commonly used in data management planning through data integration. In Chap. 2, Michener (2017a) provides guidance on developing effective data management plans. Chapter 3 (Porter 2017) describes different database approaches that can be used to organize and manage data, as well as key data management concepts like data modelling and data normalization. Chapter 4 (Michener 2017b) focuses on commonly used graphical and statistical QA/QC approaches to ensuring data quality. In Chap. 5, Michener (2017c) discusses the metadata standards and tools that can be used to document data so it can be easily discovered, accessed and interpreted. Cook et al. (2017) describe best practices for protecting and preserving data to support long-term acquisition and use in Chap. 6. Chapter 7 (Michener 2017d) focuses on methods that can be employed to more easily discover data as well as make data more readily discoverable by others. In Chap. 8, Schildhauer (2017) discusses the underlying principles and practices involved in integrating data from different sources—a necessary prerequisite for most data analysis and synthesis efforts.

### 1.3 Analysis and Synthesis

The five chapters in Part III discuss a subset of modern tools that can be used for analysis, synthesis and forecasting. Figure 1.4 provides an overview of basic steps and methods of data analysis and synthesis in ecology. *Conceptual Models* should be the starting point by reflecting research questions and key variables in an instructive way. Sources for *Data Acquisition* typically include field, laboratory and/or literature data. Common methods for *Data Analysis* are canonical correspondence analysis (CCA), principal component analysis (PCA) as well as self-organising maps (SOM) that reduce the data dimension and reveal nonlinear relationships by ordination and clustering of multivariate data.

In Chap. 12, Park et al. (2017) address explicitly the benefits of SOM for revealing and visualising nonlinear relationships in complex ecological data, and in Chap. 18, Mihuc and Recknagel (2017) demonstrate applications of canonical correspondence analysis for qualitative analysis of interrelationships between the native zooplankton community and invasive zebra mussel and alewife in Lake Champlain.

*Data Synthesis* can be performed by statistical, inferential and process-based modelling techniques. Statistical modelling basically utilises univariate nonlinear and multivariate linear regression analysis but fail to identify multivariate nonlinear relationships intrinsic of ecological data. By contrast, inferential models using artificial neural networks (ANN) and evolutionary algorithms (EA) are well suited

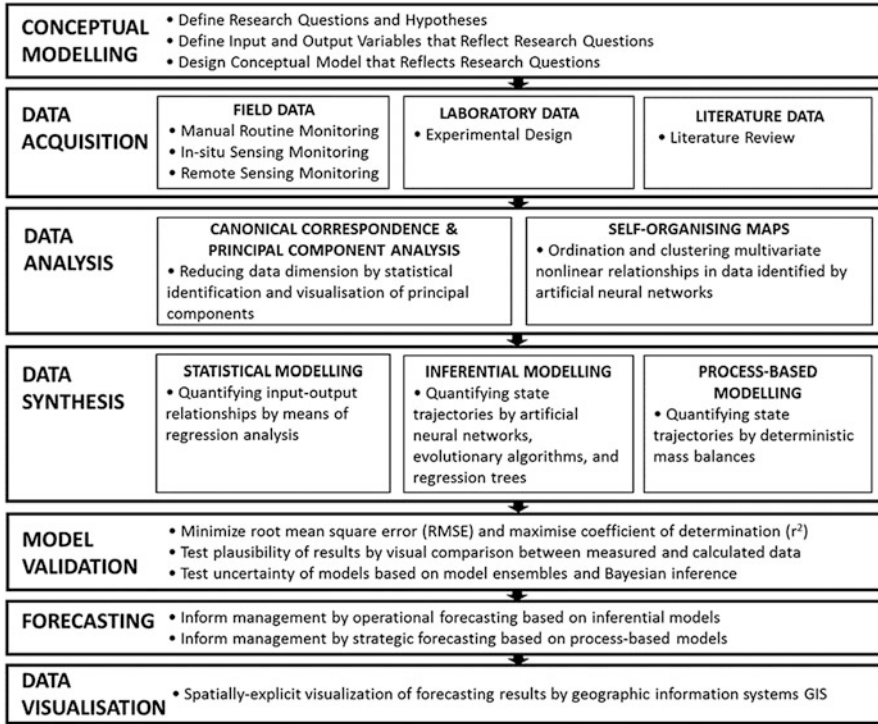


Fig. 1.4 Basic steps and methods of data analysis and synthesis

as tools to encapsulate and predict the highly complex and interrelated behaviour of ecological entities solely based on inductive reasoning. ANN do not explicitly represent models and this is viewed as a major shortcoming of this computational technique. By contrast EA represent models explicitly by IF-THEN-ELSE rules. In Chap. 9, Recknagel et al. (2017a) introduce the rationale of the hybrid evolutionary algorithm (HEA) and demonstrate applications of HEA for threshold identification, predictive modelling and meta-analysis. Inferential modelling by HEA proves also suitable for operational forecasting and early warning as discussed in Chap. 15 by Recknagel et al. (2017b). Inferential models by regression trees represent correlations between habitat properties and ecological entities by hierarchical structured IF-THEN-ELSE rules. Case studies in Chap. 15 demonstrate their capability to identify threshold conditions responsible for changing ecological entities.

Process-based models as outlined in Chap. 10 by Arhonditsis et al. (2017a) synthesize data by nonlinear differential equations that contain algebraic equations of Michaelis-Menten-type kinetics, causal and empirical relations. As demonstrated by case studies in Chap. 10, process-based modelling of specific ecosystems requires substantial data sets as well as *ad hoc* parameter optimization and calibration. If simulation results achieve reasonable validity for a specific ecosystem as indicated by a ‘low’ root mean squared error RMSE and a ‘high’ coefficient of

determination  $r^2$ , the underlying model may be applied for hypotheses testing or long-term forecasting by scenario analysis. The credibility of scenario analyses may be constrained by the scope and inherent uncertainties of models. The analysis of model uncertainty by means of Bayesian inference is explicitly addressed in Chap. 11 by Arhonditsis et al. (2017b) and demonstrated by several case studies.

*Forecasting* of ecosystem behaviour is prerequisite for preventing or mitigating events that cause rapid deterioration of ecological entities. In Chap. 16, Recknagel et al. (2017c) address forecasting by model ensembles in order to overcome single model constraints. Case studies in Chap. 16 demonstrate that ensembles of complementary models extend the scope of an individual model, which is necessary to more realistically reveal complex interrelationships between adjacent ecosystems such as catchments and lakes under the influence of global change, and that model-specific uncertainties may be compromised by Bayesian analysis of ensembles of alternative models (see also Chap. 11). As shown in Chap. 15 (Recknagel et al. 2017b), predictive inferential models and remote sensing appear capable of short-term forecasting of rapid outbreaks of population density. Two case studies demonstrate that inferential models based on HEA allow early warning of harmful algal blooms in lakes by real-time forecasts up to 30-day-ahead. The chapter also discusses the potential of remote sensing for real-time monitoring of the spatio-temporal distribution of water quality parameters and cyanobacteria blooms in water bodies. *Data Visualisation* is prerequisite to successfully communicate and disseminate findings from data analysis and synthesis. In Chap. 13, Rocchini et al. (2017) address the potential of GIS-tools to visualise spatially-explicit modelling and forecasting results.

## 1.4 Communicating and Informing Decisions

Research findings must be accessible to technical and general audiences to inform decision-making, contribute to new knowledge, and educate about complex topics. Part IV includes three chapters that illustrate how information can best be conveyed to diverse audiences. In Chap. 14, Budden and Michener (2017) discuss best practices for communicating and disseminating research outputs via publications, presentations, illustrations and social media. Various modelling approaches can be particularly useful for informing near-term and long-term decisions. In Chap. 15, Recknagel et al. (2017b) highlight the potential for inferential models and remote sensing to inform operational decisions by short-term forecasting. In Chap. 16, Recknagel et al. (2017c) present scenario analysis by complementary and alternative model ensembles that can inform strategic decision-making by long-term forecasting.

## 1.5 Case Studies

The five specific case studies included in Part V illustrate how ecological informatics has evolved to meet the needs of the various disciplines that comprise the domain of ecological science. In Chap. 17, Parr and Thessen (2017) present two user stories that highlight the latest tools and procedures that are used to manage biodiversity data, including identification tools, phylogenetic trees, ontologies, controlled vocabularies, standards, and genomics. In Chap. 18, Mihuc and Recknagel (2017) demonstrate applications of CCA and HEA to long-term limnological data of Lake Champlain (USA). In Chap. 19, Hanson et al. (2017) provide an overview of the Global Lake Ecological Observatory Network and emphasize the role of coordinated social and technical change in a successful research network. Chapter 20 (Hong et al. 2017) describes efforts to analyse and synthesize data resulting from the Nakdong River (South Korea) Long Term Ecological Research effort. Maberly et al. (2017) report research outcomes from the LTER English Lake District in Chap. 21.

## References

- Arhonditsis G, Recknagel F, Joehnk K (2017a) Process-based modeling of nutrient cycles and food-web dynamics, Chapter 10. In: Recknagel F, Michener W (eds) Ecological informatics. Data management and knowledge discovery. Springer, Heidelberg
- Arhonditsis G, Kim D-Y, Kelly N, Neumann A, Javed A (2017b) Uncertainty analysis by Bayesian inference, Chapter 11. In: Recknagel F, Michener W (eds) Ecological informatics. Data management and knowledge discovery. Springer, Heidelberg
- Budden AE, Michener WK (2017) Communicating and disseminating research findings, Chapter 14. In: Recknagel F, Michener W (eds) Ecological informatics. Data management and knowledge discovery. Springer, Heidelberg
- Cook RB, Wei Y, Hook LA, Vannan SKS, McNelis JJ (2017) Preserve: protecting data for long-term use, Chapter 6. In: Recknagel F, Michener W (eds) Ecological informatics. Data management and knowledge discovery. Springer, Heidelberg
- Hanson PC, Weathers KC, Dugan HA, Gries C (2017) The global lake ecological observatory network, Chapter 19. In: Recknagel F, Michener W (eds) Ecological informatics. Data management and knowledge discovery. Springer, Heidelberg
- Hong D-G, Jeong K-S, Kim D-K, Joo G-J (2017) Long-term ecological research in the Nakdong River: application of ecological informatics to harmful algal blooms, Chapter 20. In: Recknagel F, Michener W (eds) Ecological informatics. Data management and knowledge discovery. Springer, Heidelberg
- Maberly SC, Ciar D, Elliott JA, Jones ID, Reynolds CS, Thackeray SJ, Winfield IJ (2017) Long-term ecological research in the English Lake District: from ecological informatics to the generation of ecological knowledge, Chapter 21. In: Recknagel F, Michener W (eds) Ecological informatics. Data management and knowledge discovery. Springer, Heidelberg
- Michener WK (2017a) Project data management planning, Chapter 2. In: Recknagel F, Michener W (eds) Ecological informatics. Data management and knowledge discovery. Springer, Heidelberg

- Michener WK (2017b) Quality assurance and quality control (QA/QC), Chapter 4. In: Recknagel F, Michener W (eds) Ecological informatics. Data management and knowledge discovery. Springer, Heidelberg
- Michener WK (2017c) Creating and managing metadata, Chapter 5. In: Recknagel F, Michener W (eds) Ecological informatics. Data management and knowledge discovery. Springer, Heidelberg
- Michener WK (2017d) Data discovery, Chapter 7. In: Recknagel F, Michener W (eds) Ecological informatics. Data management and knowledge discovery. Springer, Heidelberg
- Mihuc TB, Recknagel F (2017) Lessons from bioinvasion of Lake Champlain, U.S.A., Chapter 18. In: Recknagel F, Michener W (eds) Ecological informatics. Data management and knowledge discovery. Springer, Heidelberg
- Park Y-S, Chon T-S, Bae M-J, Kim D-H, Lek S (2017) Multivariate data analysis by means of self-organizing maps, Chapter 12. In: Recknagel F, Michener W (eds) Ecological informatics. Data management and knowledge discovery. Springer, Heidelberg
- Parr CS, Thessen AE (2017) Biodiversity informatics, Chapter 17. In: Recknagel F, Michener W (eds) Ecological informatics. Data management and knowledge discovery. Springer, Heidelberg
- Porter JH (2017) Scientific databases for environmental research, Chapter 3. In: Recknagel F, Michener W (eds) Ecological informatics. Data management and knowledge discovery. Springer, Heidelberg
- Recknagel F (2017) Ecological informatics. In: Gibson D (ed) Oxford bibliographies in ecology. Oxford University Press, New York. <http://www.oxfordbibliographies.com/view/document/obo-9780199830060/obo-9780199830060-0174.xml>
- Recknagel F, Kocev D, Cao H, Branco CC, Minoti R, Dzeroski S (2017a) Inferential modelling of population dynamics, Chapter 9. In: Recknagel F, Michener W (eds) Ecological informatics. Data management and knowledge discovery. Springer, Heidelberg
- Recknagel F, Orr P, Swanepoel A, Joehnk K, Anstee J (2017b) Operational forecasting in ecology by inferential models and remote sensing, Chapter 15. In: Recknagel F, Michener W (eds) Ecological informatics. Data management and knowledge discovery. Springer, Heidelberg
- Recknagel F, Arhonditsis G, Kim D-K, Nguyen HH (2017c) Strategic forecasting in ecology by inferential and process-based models, Chapter 16. In: Recknagel F, Michener W (eds) Ecological informatics. Data management and knowledge discovery. Springer, Heidelberg
- Rocchini D, Garzon-Lopez CX, Barbosa AM, Delucchi L, Olandi JE, Marcantonio M, Bastin L, Wegmann M (2017) GIS-based data synthesis and visualization, Chapter 13. In: Recknagel F, Michener W (eds) Ecological informatics. Data management and knowledge discovery. Springer, Heidelberg
- Schildhauer M (2017) Data integration: principles and practice, Chapter 8. In: Recknagel F, Michener W (eds) Ecological informatics. Data management and knowledge discovery. Springer, Heidelberg