# mangal – making complex ecological network analysis simpler

T. Poisot

Jan. 2014

This is a *working document* describing mangal, a set of JSON objects templates to encode ecological networks of virtually any complexity. There are plans to host a pilot database.

```
## Error: impossible de trouver la fonction "as"
```

## Introduction

{edit}Ecological networks enable ecologists to accommodate the complexity of natural communities, and to discover mechanisms contributing to their persistence, stability, resilience, and functioning [1, 2]. Yet, meta-analyses of a large number of ecological networks are still extremely rare, and most of the studies comparing several networks {ref} do so within the limit of particular systems. Networks, as they encode the structure of complex ecological interactions, have been time and again presented as useful tools to understand ecosystem properties and dynamics {ref}. Coming up with a clear conceptual and mechanistic understanding of the relationships between the structure of ecological networks and ecosystem properties require to pool a large quantity of data.

On the other hand, the recent years saw the development of the idea that network structure is itself a dynamical object, which will change as a function of environmental conditions and as a result of meta-community processes {ref}. Although the *existence* of this variation has been demonstrated, the reasons for which it happens are much less clearly understood, and will require a change in focus, from species to populations {ref}. Because the variability of interactions involve a host of ecological mechanisms, it is likely that important data mining efforts will be required to fully understand it. Notably, new approaches based on the replication of networks over temporal, spatial, and environmental gradients are promising, but require to have a data structure ready to accomodate the results they will produce. Beyond just describing the structure of interactions, these data will need to include informations about environmental context, population characteristics, and other relevant additional explanatory variables.

In this paper, we (i) establish the need of a data specification serving as a *lingua franca* among network ecologists, (ii) describe this data specification. Finally, we (iii) describe `mangal`, a `R` package and compagnon database, relying on this data specification. We provide some use cases showing how this new approach makes complex analyzes simpler, and allows for the integration of new tools to manipulate biodiversity resources.

# Why do we need a data specification?

Ecological networks are (often) stored as their *adjacency matrix* (or as the quantitative link matrix), that is a series of `0` and `1` indicating, respectively, the absence and presence of an interaction. This format is extremely convenient for *use* (as most network analysis packages, *e.g.* `bipartite`, `betalink`, `foodweb`, require data to be presented this way), but is extremely inefficient at storing *meta-data*. In most cases, an adjacency matrix will inform on the identity of species (in cases where rows and columns headers are present), and the presence or absence of interactions. If other data about the environment (*e.g.* where the network wassampled) or the species (*e.g.* the population size, trait distribution, or other observations) are available, they are most either given in other files, or as accompanying text. In both cases, making a programmatic link between interaction data and relevant meta-data is difficult and error-prone.

By contrast, a data specification provides a common language for network ecologists to interact, and ensure that, regardless of their source, data can be used in a shared workflow. Most importantly, a data specification describes how data are *exchanged*. Each group retains the ability to store the data in the format that is most convenient for in-house use, and only needs to provide export options (*e.g.* through an API) respecting the data specification. This approach ensures that *all* data can be used in meta-analyses, and will in time increase the impact of data {ref}.

# Elements of the data specification

{complete}The data specification is built around the idea that (ecological) networks are collections of relationships between ecological objects, each element having particular meta-data associated. In this section, we detail {complete}. An interactive webpage with the elements of the data specification can be found online at `http://mangal.uqar.ca./doc/spec/`. The data specification is implemented as a series of `JSON` schemes, *i.e.* documents describing how the data should be formatted, and what each element represent. The schemes can be downloaded from `https://github.com/mangal-wg/mangal-schemes/releases/tag/1.0`.

Rather than giving an exhaustive list of the data specification (which is available online at the aforementionned URL), this section will propose an overview of each element, and of how they interact. Within the `R` package, information about the data format can be viewed using the `whatIs` function (*e.g.* `whatIs(api, 'taxa')` will return a table with information about how `taxa` objects are formated.

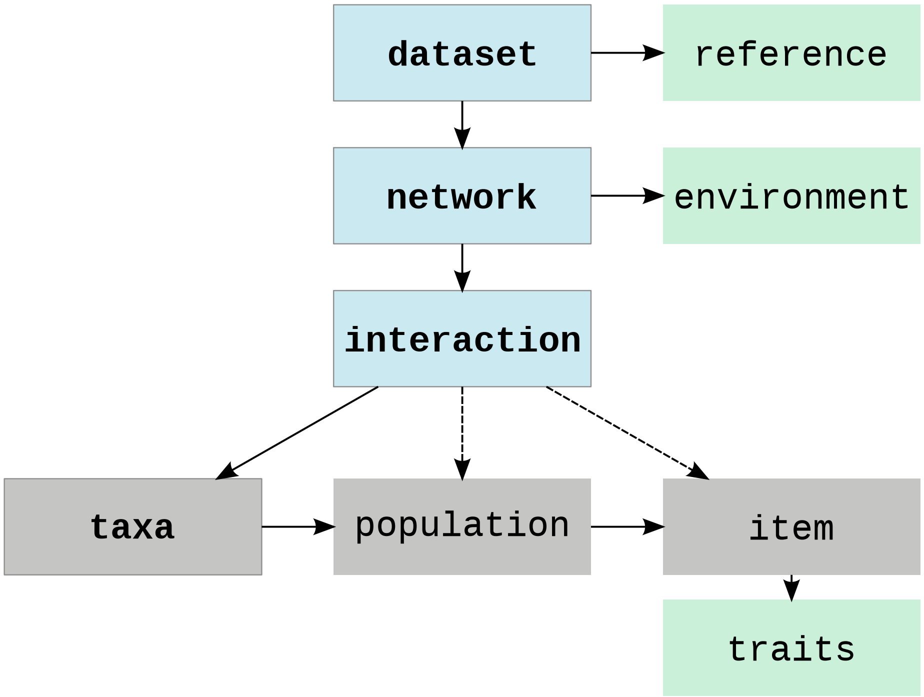

Figure 1: An overview of the data specification, and the hierarchy between objects. Each box correspond to a level of the data specification. Grey boxes are nodes, blue boxes are interactions and networks, and green boxes are metadata. The **bold** boxes (`dataset`, `network`, `interaction`, `taxa`) are the minimal elements needed to represent a network.

We propose `JSON` as the most efficient data format for the following reasons. First, it has emerged as a *de facto* standard for web platform serving data, and accepting data from users. Second, it allows *validation* of the data: a `JSON` file can be matched against a scheme, and one can verify that it is correctly formatted. Finally, `JSON` objects are easily and cheaply (memory-wise) parsed in the most common programming languages, notably `R` (equivalent to `list`) and `python` (equivalent to `dict`).

## Node informations

### Taxa

Taxa are a taxonomic entity of any level, identified by their name, vernacular name, and their identifiers in a variety of taxonomic services. Associating the identifiers of each taxa is important to leverage the power of the new generation of open data tools, such as `taxize` {ref}. For example, a taxa with an associated *NCBI Taxonomy* identifier can be represented this way:

```
{
    "name": "Lamellodiscus ignoratus",
    "vernacular": "Lamellodiscus ignoratus",
    "ncbi": "142934"
}
```

The data specification currently accomodates `ncbi`, `gbif`, `itis` and `bold` identifiers. Correspondances between these and other services can be made through other tools, such as *e.g.* `taxize`.

### Population

### Item

## Network informations

### Interaction

### Network

### Dataset

## Meta-data

### Trait value

### Environmental condition

### User

paternity {ref}

**References**

# Use cases

{edit}In this section, we present use cases using the `rmangal` package for `R`, to interact with a database implementing this data specification, and serving data through a `REST`ful API (`http://mangal.uqar.ca/api/v1/`). It is possible for users to deposit data into this database, through the `R` package. Data are made available under a *CC-0 Waiver*.

```
library(rmangal)
api <- mangalapi()
```

## Plotting a network

```
graph <- network_as_graph(api, 2)
plot(graph)
```

## Network beta-diversity

## Connectance and richness relationships

# References

1. Dunne JA: **The Network Structure of Food Webs**. In *Ecological networks: Linking structure and dynamics*. edited by Dunne JA, Pascual M Oxford University Press; 2006:27–86.

2. Blüthgen N, Fründ J, Vázquez DP, Menzel F: **What do interaction network metrics tell us about specialization and biological traits?** *Ecology* 2008, **89**:3387–99.

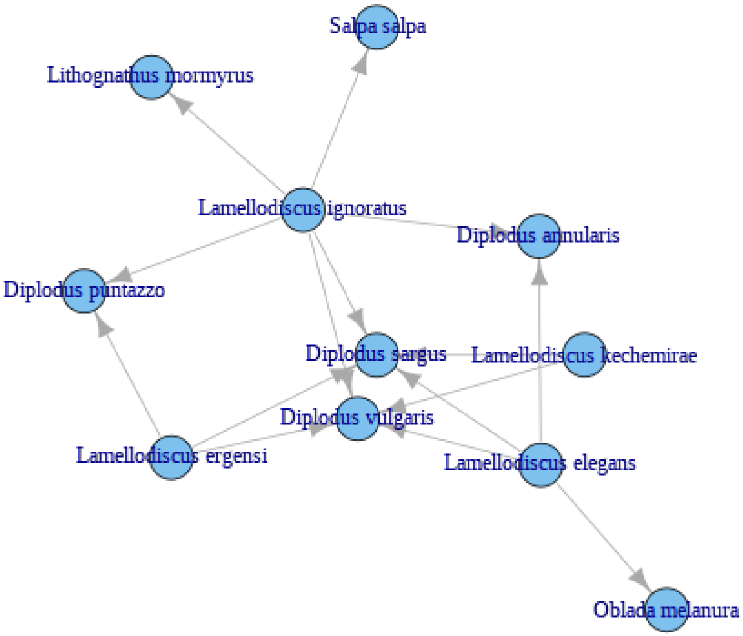

Figure 2: Example of network plotting, using the `network_as_graph` function.