



From rasters to gradients of a social ecological research network

Isabelle Charpentier

► To cite this version:

Isabelle Charpentier. From rasters to gradients of a social ecological research network. 2023. hal-03942291v2

HAL Id: hal-03942291

<https://hal.science/hal-03942291v2>

Preprint submitted on 9 Feb 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

From rasters to gradients of a social ecological research network

Isabelle Charpentier^{a,b}

^aLTSE-FR Zone Atelier Environnementale Urbaine, 5 allée du Général Rouvillois, 67000 Strasbourg, France

^bICube, UMR 7357, Université de Strasbourg and CNRS, 67412 Illkirch, France

Abstract

Open, world wide, geographical information databases are frequently mobilized for hydrological studies in poorly gauged watersheds. More generally, the variety of such databases – landscape, climate, population, vegetation, land cover, practices and so forth – is an opportunity for sharing insights amongst a social ecological research network or a federation of protected areas. In practice, dealing with the heterogeneity in the variables, in the datasets and in the providers, remains a challenge as it often raises some of well-known interoperability and reuse issues.

Through software engineering, some general workflow is set up from raster extraction to the computation of social and ecological indicators. The heterogeneity in the data allows for a reflection on FAIR practices amongst a number of providers on the one hand, and the fairification of the products of this research in the other hand.

As a case study, the workflow is operated on the French long term socio-ecological research network, namely Réseau des Zones Ateliers, the nodes of which conduct action-oriented studies on social ecological systems (SESs) such as river basins, rural, urban or mountainous territories, or protected areas in Metropolitan France and outside Europe. As expected, some gradients reflecting the diversity in the territories already appear in a half-page pie chart table representing computed indicators.

Keywords: Geoinformatics, social ecological system, satellite-based raster, FAIR, indicator.

1. Introduction

Open, world wide, geographical information databases are an opportunity for social ecological research networks or federations of protected areas. Through software engineering, some general workflow may be set up from raster extraction to the computation of social and ecological indicators. Although the raster accuracy is a limitation at the level of fine physical processes, gridded satellite-based data are of interest for partly monitored areas, as demonstrated for small to large poorly gauged watersheds. As an example, the very basic gridded variables – climate, population, vegetation and land cover – may already provide a coarse but relevant information on a set of social ecological systems.

On the one hand, retrieving and compiling heterogeneous data is among the scientific hot issues. For instance, in Europe, the Long-Term Ecosystem Research network (eLTER) maintains the Dynamic Ecological Information Management System - Site and dataset registry (DEIMS-SDR) (Wohner et al., 2019). eLTER also develops an R package called ReLTER (Oggioni et al., 2021) for data extraction from publicly available re-

sources in order to enrich on-site observations. On the other hand, publicly available data are not always open enough, neither to a general audience nor to a computer workflow. This results from a lack of information on data management (content, format, unit, protocol, and so forth) that hampers both interoperability and reusability.

To face this issue, the Findable–Accessible–Interoperable–Reusable guiding principles (FAIR) were proposed for scientific data management and stewardship (Wilkinson et al., 2016). Although the FAIR philosophy is adopted by a number of scientific communities, the FAIR implementation remains a challenge in a social ecological research due to the heterogeneity of variables and sources. This is even more true at the level of a research network, where a minimum of generality in the protocols and the data may facilitate fair comparisons, notably along gradients.

The main goal of the paper is to provide a quantitative insight of the gradients of a social ecological research network. To that end, a general workflow is designed. In a nutshell, world wide rasters are reworked

Variable	Abbr.	SRID	Resol.	Years	Units	Format	Provider	Product
Temperatures	LSTD	4326	0.05°	2001–20	°F	HDF	USGS	MOD11C3
	LSTN	4326	0.05°	2001–20	°F	HDF	USGS	MOD11C3
Vegetation	NDVI	4326	0.05°	2001–20	–	HDF	USGS	MOD13C2
Biogeoregions	BIOG	3035	(1:1M)	2015	12 classes	shp	EEA	Data
Pop. density	GPW4	4326	30"	2020	pers./km ²	TIF	SEDAC	GPW, v4
Land cover	LCxx	4326	10 m	2016	10 classes	TIF	ESA	World Cover 2020 v100
ZA shapes	–	4326	–	2022	°	JSON	eLTER	DEIMS-SDR
World Countries	–	4326	–	2020	°	shp	efrainmaps.es	(Porto Tapiquén, 2020)
Colormaps	batflow	–	8 bits	2018	RGB	txt	Zenodo	(Cramer, 2018)
	turbo	–	8 bits	2019	RGB	txt	Matlab	

Table 1: Variables and datasets mobilized in the study of social ecological gradients.

and cropped with respect to the territories of interest so as to compute averages or other zonal indicators. As expected, the reuse of open datasets is not an easy task. Issues are discussed from the FAIRness of the chosen data and to the FAIRification of the workflow sub-raster and indicator products. A special attention is brought to the graphical rendering of the indicators through pie charts.

The French long term socio-ecological research network, namely Réseau des Zones Ateliers (RZA) (Bretagnolle et al., 2019), is considered as a case study. This distributed inter-agency research network is affiliated to the eLTER network. Fifteen Zones Ateliers (ZAs) conduct action-oriented studies on essential social ecological systems (SESSs) such as river basins, rural, urban or mountainous territories, or protected areas (see Fig. 2). Although the network has started rolling out open data, the heterogeneity in the observed data, in their acquisition protocols and in their delivering formats restrains a quantitative assessment of ecological and land use gradients.

The layout of the paper is as follows. Some definitions and constraints about reused data and foreseen products are presented in Section 2. The corresponding workflow is discussed in Section 3 with respect to FAIRness and FAIRification points of view. Resulting zonal indicators, their meaning and their accuracy are reported and analyzed in Section 4. Section 5 provides a conclusion and raises some research and open science opportunities.

2. Data and products

Today, world wide temporal sequences of climate, land use and land cover, vegetation and population rasters are made available by national and European agencies giving information about very basic social ecological variables. That being said, questions arise about findability, accessibility, interoperability and reusability of the data, but also in terms of accuracy, heterogeneity, and representation. For replicability, we consider findable and accessible data only.

This section recalls general features about the grid/raster data and shapefiles listed in Tab. 1, details being exposed in section 4. Methods for indicators are then presented.

5 km at the Equator. A monthly frequency is chosen for the climate and vegetation variables. Image processing of raw satellite pictures was carried out by producers prior to the distribution of their datasets.

Table 1 reports some variables (names and abbreviations) and key dataset characteristics (SRID, spatial resolution, availability ranges, units, formats, providers, product names). The heterogeneity observed in Tab. 1 is of interest to audit the FAIR principles and to evaluate various social ecological indicators.

Such datasets may be managed either by a geographical information system software like QGIS or by a scientific computing programming language like Matlab, Python or R.

2.1.1. Grids

A number of world wide grids are gathered on the NASA Earth Observations (NEO) website into atmosphere, energy, land, life and ocean themes. These rasters may be browsed and downloaded at a number of low resolutions as JPEG images for a general audience or as TIFF files. Some links to the data producers' web portals are provided to the readers interested in higher resolution datasets.

Data providers deliver rasters as text files (ASCII), images (JPEG, TIFF), as images with geographical metadata (GEOTIFF), or as structured sets of grids (HDF and NETCDF). Their websites are sometimes designed to allow for selective download (area, spatial resolution, period, variables), possibly through standard web protocols like Web Map Service (WMS) or Web Feature Service (WFS). The File Transfer Protocol (FTP) may be pertinent to a research at a world wide scale while avoiding multiple and repetitive accesses to very large datasets.

Datasets comprising a bundle of grids are stored using either the Hierarchical Data Format (.hdf files) or the Network Common Data Form (.nc files). Related files then require some preliminary work to extract the

grids of interest. Unpacking these files is not always an easy task, even though some procedures and piece of code are proposed (Earth Lab).

2.1.2. JSON files

At the time of writing, DEIMS-SDR (Wohner et al., 2019) is the most comprehensive public catalogue describing the ZAs and the other European long-term ecosystem research sites.

Information on the eLTER sites was collected on a common declarative basis and stored into individual structured JSON files. In short, JavaScript Object Notation is plain text encoding data as name/value pairs and organized into a hierarchical structure. This hierarchical textual structure may be viewed in a browser.

Among attributes, we use the geographic boundaries to reconstruct the polygons of the ZAs only. Note that DEIMS-SDR also provides the shapefiles that can be downloaded individually from the website. In that case, their download link may be reconstructed from data in their JSON files.

2.1.3. Shapefiles

A shapefile is a set of files (.shp, .shx, and .dbf) encoding geographic information such as the geometry (points, lines or polygons) that delineate a zone, some indices and attributes features, respectively. A shapefile preferably comes with a .prj file informing on the projection and the coordinate system. This is the case of the biogeographical regions. Shapefiles are used to crop rasters in order to evaluate zonal indicators.

2.2. Methods and outcomes

Main computational steps combining rasters and shapefiles for the computation of indicators are briefly presented in this subsection. A comprehensive workflow is presented in section 3 with regards to FAIR perspectives. The meaning and accuracy of the indicators are discussed in section 4.

2.2.1. Programming environment

The geographical information system (GIS) software ArcGIS and QGIS provide graphical user interfaces and programming interfaces with Python. In contrast, Topo-toolbox (Schwanghart and Kuhn, 2010) is a set of Matlab functions dedicated to digital elevation model analysis, with a special focus on hydrology. It can handle geographical rasters.

The comparison of performances reported in (Schwanghart and Scherler, 2014) for raster handling, the scientific computing and graphical user

interface abilities provided by the Matlab framework are decisive factors in selecting Matlab as a programming environment. In particular, this framework was effective in the design of complex partition methods into area-balanced sub-watersheds (Hariri et al., 2022) and into some hexagonal partitions for serious game boards (Charpentier, 2021).

In the absence of a specific mention, the computations are carried out using TopoToolbox and the Matlab programming language. Computations are carried out on a small laptop (11th Gen Intel® Core™ i7-1165G7 @ 2.80GHz×8” with 16 GiB of RAM memory) without a specific use of the Matlab parallel pool.

2.2.2. Crop process

Any raster may be cropped using a shapefile. This yields a smaller raster (matrix representation), with default numerical values or NaN values outside the zone of interest, and a vector of indices indicating the raster cells/pixels within the zone. The boundary cells depend on the cropping method.

Geometries such as points or very small polygons are viewed as a unique index/pixel of the raster, corresponding to a unique cell with an area of about 5×5 km² in our case. Lines or tight polygons may lead to a very few pixels too.

2.2.3. Averages and normals

The Guide to Climatological Practices (WMO) recalls a number of useful terms corresponding to different periods. These terms are suitable for long term socio-ecological studies and are used to define indicators.

Averages are computed as monthly means (temperature, for instance) or monthly totals (precipitation, for instance). *Period averages* are computed on periods of at least ten years starting on 1st January of a year ending with the digit 1. *Normals* involve at least three consecutive ten-year periods. These are not yet available from satellites rasters (see ranges in Tab 1). Where applicable, these definitions may be extended to any timeseries.

2.2.4. Pie charts

Thinking in the representation of social ecological gradients, we choose pie charts to gather a maximum of heterogeneous information in a minimum of space. As displayed in section 4, pie charts may account for temporal (MODIS data) or spatial (population density, GPW4) variations.

Heterogeneity in the variables may be tackled through the use of a unique color map for temperature,

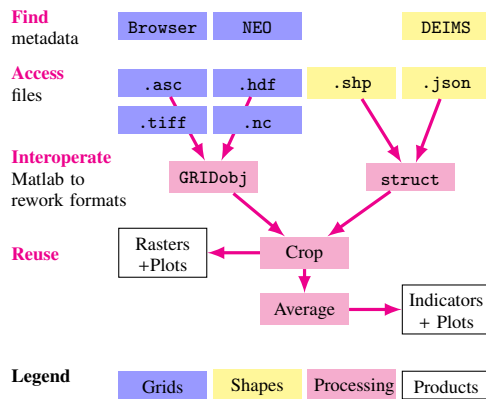


Figure 1: Workflow from a reuser point of view.

vegetation and population. For instance, we choose the batlow color map for NDVI and population density. For the latter, the green tones stand for green vegetated or rural areas, while the pink tones are related to bare, built up or urban areas. We prefer the use of the turbo color map with blue to red tones for temperatures.

3. The two sides of FAIR

FAIR – Findable, Accessible, Interoperable and Reusable – principles (Wilkinson et al., 2016) address digital data and metadata as well as data infrastructure aspects. These may give the impression of repeats. In other words, when a dataset is not FAIR, it usually fails on several of the four criteria.

Thinking in open science, this section unravels and revisits the FAIR data principles from a reuse point of view to go one step beyond their description. As FAIRification is not an easy task, we provide a firsthand account of the Find, Access, Interoperate, and Reuse procedures carried out during the research.

The discussion is supported by the workflow (Fig. 1) implemented to convert bio- and geophysical rasters into social ecological indicators. This ranges from the digital data collection (subsection 3.1), passing through the transformation into new data and knowledge (subsection 3.2), to the fairification of the products. FAIRification and FAIR-checkers are discussed in subsection 3.3.

3.1.1. Finding the data

Findable data are described through metadata (FAIR or not) accessible through a browser, somewhere on the Internet.

Data are searched with a particular objective, using a few key words related to basic social ecological variables (climate, earth, temperature, population), sensing devices (satellite or ground based), data properties (rasters, tables). Strings such as “Climate raster data”, “vegetation index raster”, “land cover maps” work well to reach US and European agencies’ data portals. “Web services” or “Download” may be added to the query. The success of such queries depends on the metadata supporting the web-pages, the papers or the datasets of interest.

3.1.2. Accessing the data

Governmental or European organizations are major providers of rasters and shape data following recommendations on Open Science. Scientific publications may also provide links to data.

Data portal accesses are granted in very different manners, with protocols ranging from web-services and fluxes that ensure interoperability with GIS software, to File Transfer Protocol (FTP). A user authentication is often required.

Providers may bundle their rasters. This is the case for the MODIS HDF files (see Tab. 1) that contain more than 10 grid variables each. The required memory may be much larger than expected when the data are unpacked. On-line cropping tools and multiple resolutions are often proposed to limit the bandwidth and memory consumption. A unique download may be preferable to repeated WMS or WFS accesses.

3.1.3. Interoperating the data

Interoperability is a computer system property that implies exchanges between one or more computer systems or products belonging to different providers by using open standards. Where data are accessible, additional meaningful and formatted metadata are necessary to ensure interoperability. As an example, interoperability is provided by GIS software when datasets following geospatial standards (ESRI) are reachable through fluxes. Interoperability thus depends on data producers’ and providers’ practices.

In Fig. 1, the Interoperate step is not Interoperability since some of dataset formats (HDF and NETCDF) are not following open standards. Two technical options arise to manage particular files and formats. On the one hand, the missing interfaces may be implemented by the

user in the GIS tool. On the other hand, the datasets may be downloaded and reworked (format, resolution, scaling, documentation) in a local repository. The second option is implemented.

3.1.4. Reusing the data

Users (citizens, researchers, policy makers) looking for data have different background and different objectives. The R of FAIR does require the availability of data (F and A principles) and software (I principle) making them accessible and affordable.

Today, more and more web portals provide data visualization and some data processing and analysis. Developer groups are targeted by initiatives such as the Earthdata Developer Portal to build and feed their own portals. Corresponding applications search, access, and interoperate within other data portals so as to propose new products. The R package called ReLTER (Oggioni et al., 2021) is built that way.

In our case, FAIRness is not always consistently met. We choose to download rasters once so as to rework the data (unpacking, reprojecting to WGS 84, resampling at a resolution of 0.05°) before their actual reuse in the computation of the indicators.

3.2. From data to indicators

At this stage of the workflow, world wide formatted scaled rasters are on the computer, ready to be cropped. The challenge of the workflow is in the preparation of heterogeneous data, the management of the multiple files and their FAIR publication in a data warehouse, preferably carried out in an automated manner.

Three nested loops operating on (1) the list of the sites, (2) the list of 15 variables (NDVI, LSTD, LSTN, GPW4, BIOG and 10 land cover classes), (3) the temporal range (20 years \times 12 months for TERRA/MODIS variables) yield a total number of 729 sub-rasters per site (3 months are missing for NDVI). These are spread into site repositories.

Averages and quantiles are computed for each site. The temporal series (temperature and vegetation) are saved as CSV spreadsheets. Heterogeneous indicators are gathered into a hierarchical structure and exported as JSON and XML files to facilitate interoperability. These new outcomes may enrich the JSON files describing the sites and stored on DEIMS-SDR.

As a cross-validation, indicators are plotted as pie charts by reading the new JSON files, see Fig. 3 in section 4.

3.3. FAIRification of the products

When available, the data management plan (DMP) of a project or a research unit contains the information about the life cycle of the data, from the collection and creation to the sharing and archiving stages, and the resource people (data protection officer and data steward, for instance).

From a researcher perspective, the post-production data management should be delegated to some disciplinary research infrastructure or some more general data warehouse so as to benefit from their FAIRification procedures.

The public data warehouse called InDoRES (Inventaires des Données en Environnement et Sociétés) is targeted.

As the key goal of FAIR is to facilitate the reuse of data, our implementation of the FAIR principles (see labels of the FAIR principles in subsections 3.3.1 to sec:findable) is discussed in reverse order, that is from Reusable to Findable.

3.3.1. Reusable

Reuse is facilitated when the FAIR workflow (metadata, data, infrastructure) is described exhaustively and in meaningful terms (R1). Metadata include a license (R1.1), the information on the origin of the data (producer, protocol, formats, vocabulary (Parsons et al., 2022)...) (R1.2). Community standard formats and vocabulary are preferred for metadata and data (R1.3).

For each variable, the metadata listed above are gathered into a structure implemented in a Matlab function, then exported as JSON and XML files. The license information is compiled from these JSON files. The naming convention is <variable>.json and license.json. Site repositories contain copies of these files.

For each site, the indicator file <site>.json is structured with respect to the variables. It also contains the simplified shapefile used for the crops. This information is exported as JSON and XML files. A merge with the DEIMS-SDR JSON file could be performed. Sub-rasters are saved as ASCII files named <site>-<variable>-<yyyy>-<mm>.asc, using the same SRID (WGS 84) and the same resolution. Timeseries are saved as <site>-<variable>.csv. Any change is propagated by running the computer code again.

Reusability is mainly a data producer task, supported by data stewardship provided by the infrastructure if necessary.

3.3.2. Interoperable

Interoperability principles emphasize best practices on knowledge-representation (I1) related to a specifica-

tion language (XML), notations (JSON), vocabularies (I2) and other qualified references (I3).

The generated datasets (about 740 files per site including metadata, data and license) are deposited site by site on the InDoRES Dataverse, together with corresponding metadata sheets. A general metadata sheet compiling information at the level of the research network is deposited as a parent sheet. Preferred vocabularies are AGROVOC, GEMET and ENVTHES.

Interoperability is an infrastructure task (templates for metadata sheets, lecture on best practices) and a data producer task.

3.3.3. Accessible

Accessibility assumes a dedicated infrastructure, managed following FAIR principles. Standardized communications protocol (A1), open protocols (A1.1), authentication (A1.2), long term availability of the metadata (A2). Accessibility is an infrastructure task.

Three filing processes are tested. Firstly, the InDoRES website proposes an Excel sheet and guides to fill metadata. These are then published by an administrator of the InDoRES Dataverse. Obviously, an automated procedure would be more convenient for the publication of metadata sheets related to the sites of a research network.

Secondly, Geoflow (Blondel et al., 2020) is an R engine allowing to run geospatial (meta)data workflows on infrastructures hosting either a GeoNetwork catalogue, a GeoServer, a Dataverse, or on Zenodo. The data producer interacts with Geoflow through a formatted CSV sheet, by using as many lines as datasets. Geoflow may be well suited to our case study since the sites' metadata differ on their names and spatial extension only. A small code suffices to generate such a CSV file. This test of Geoflow is successful on two GeoNetwork and GeoServer sandboxes.

Finally, the add button is available to people with the admin rights. Metadata are then filled using the Dataverse interface.

At the time of writing, the parent dataset is published together with data containing the indicators computed for the different sites (Charpentier, 2023). Geoflow is planned for the deposit of the raster files, using one metadata sheet and one zipped dataset per site.

3.3.4. Findable

Findability is managed at the infrastructure level for the provision of both digital object identifiers (doi, F1 and F3) and a searchable resource (F4). The producer task (F2) consists in the provision of a dataset together with rich metadata.

Metadata met the InDoRES requisites in terms of FAIR principles when using the Excel sheet. Geoflow sheets are based on Dublin Core. Some of the metadata are described in subsection 3.3.1.

3.3.5. FAIR-checkers

The FAIRification of the products is conducted with the hope that data will be found, accessed and reused through an interoperable system. Some FAIR-checkers (Hernandez Serrano and Emonet, 2022; Rosnet et al., 2020) are being developed to evaluate the degree of FAIRness of a dataset. This may help the data producer in the choice of an infrastructure too. However, there is still a lot of work to be done around FAIR metrics before getting reliable results (Wilkinson et al., 2022).

The parent dataset fulfills 16/17 principles using Fairchecker (Rosnet et al., 2020) and reaches a moderate level of FAIR (52%) using F-UJI (Hernandez Serrano and Emonet, 2022) when using the digital object identifier of the dataset deposited on the InDoRES' Dataverse. In comparison, the metadata gathered for the biogeoregions (European Environment Agency, 2016) fulfills 15/17 criterions (when using the link) because the metadata do not include persistent identifier. These metadata also reach a moderate level of FAIR (60%) while using F-UJI.

4. Results

Several open datasets (raster and shapes, see Tab. 1) provided and found on the web with sufficient information are considered for the design of social ecological system indicators at the level of a research network. These first indicators are intended to support studies along some of the so-called social ecological gradients (landscape, urbanization, climate, elevation, practices, North-South, water availability...).

The French long term social ecological research network is considered as a case study for the diversity of its SESs (subsection 4.1). The general workflow is specified for each dataset in subsection 4.2. Results on indicators and gradients are presented in subsection 4.3. Finally subsection 4.4 discusses the workflow abilities as well as some of the limits of the proposed set of indicators.

4.1. Case study

The Réseau des Zones Ateliers (RZA) is a social ecological research network with an emphasis on human-nature interactions and sustainability.

The social ecological systems are an alpine SES (ZAA, Alpes), an urban SES (ZAEU, Urban Environment), agricultural SESs (ZAAJ, Arc Jurassien; ZAAR, Armorique; ZARG, Argonne; ZAPVS, Plaine et Val de Sèvres), SESs related to water bodies (ZABR, Rhone river; ZABRI, Brest-Iroise; ZAL, Loire river; ZAM, Moselle river; ZAPYGAR, Pyrénées-Garonne; ZA-Seine, Seine river), SESs in the southern hemisphere (ZAHWG, Hwangé natural park, Zimbabwe; ZATA, French Antarctic and Southern Lands), and a SES with past uranium mining (ZATU). Some of the Metropolitan France zones under study are represented in Fig. 2. The new ZARG is not on DEIMS-SDR for the moment.

The network promotes the long term observation of environmental practices by considering a mix of social and biogeophysical variables through on-site experiments. Accordingly, the ZA teams bring together ecological sciences and humanities, researchers, citizens and local authorities. The interested reader may find a computational assessment of the inter- and trans-disciplinary nature of the RZA's research through natural language processing in (Falk and Charpentier, 2021).

4.2. Datasets

Rasters for land surface temperature (LST), normalized difference vegetation index (NDVI) and gridded population of the world (GPW4) were reached by means of the NEO site. The biogeoregions shapefile and the global land cover (LC) raster were discovered using Google. Their reading requires more or less work, depending on the file format, the resolution and the coordinate system. Note that precipitations are not considered since data from tropical rainfall measuring mission (TRMM) are not accurate at European latitudes.

Data were downloaded and processed as indicated in the workflow to extract and/or resample information from WGS 84 rasters with a unique resolution of 0.05°. Rasters are first cropped to $[-6^{\circ}, 9^{\circ}] \times [41, 52]$, a bounding box that comprises France, to reduce the time consumption for further calculations. While a total amount of 70.8 GiB are downloaded (world wide raster for MODIS and GPW4, and 3 tiles for ESA-GLC representing 32.7 GiB alone), cropped rasters represent an amount of 0.10 GiB. JSON files were downloaded from DEIMS-SDR in September 2022.

Figure 2 plots the sub-rasters cropped using the ZAs' polygons. The difference in the areas and the shapes is to be accounted in the analysis of the indicators based on a spatial average. Social ecological interpretations are discussed in subsections 4.3 and 4.4.

Some of the workflow tasks are specified for each variable in the following paragraphs as an assessment

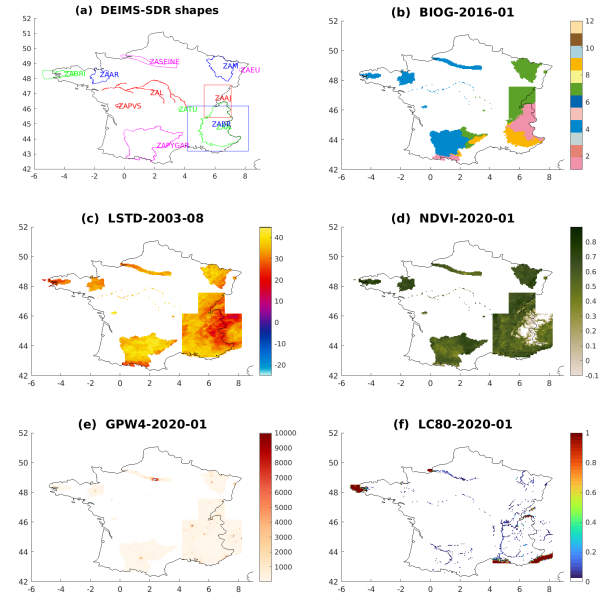


Figure 2: Raster extractions for the Metropolitan Zones Ateliers. a: Polygons stored in DEIMS-SDR JSON files, b: Mix of biogeoregions, c: Land surface temperature (heat wave of August 2003), d: Normalized difference vegetation index, e: Population density, f: Hydrosystem land cover class.

of computer and social ecological heterogeneities.

4.2.1. Land surface temperatures

The MOD11C3 Version 6 product acquired from the TERRA/MODIS satellite (Wan et al., 2015) is provided as a HDF files encoding 17 grids at a resolution of 0.05°, with a unit of 0.02 °F. Unpacking these HDF files for the first time is tricky.

We consider Land Surface Temperatures 'LST_Day_CMG' and 'LST_Night_CMG', renamed as LSTD and LSTN, at a monthly frequency in the 2001–2020 period. The data are used as is, that is the quality information is not used. This represents 240 grids for the variables LSTD (plus 240 grids for LSTN).

For each of the ZAs, the 240 files are cropped with respect to its polygon, generating 240 small sub-rasters, rescaled using a unit of 0.01°C.

The month spatial mean is computed by averaging on terrestrial temperature values only. The period monthly average computed over the 20 years and the yearly means are saved in the LSTD field of the ZAs' JSON file. The process is repeated for the LSTN variable.

Note that sea surface temperatures are available in AQUA/MODIS datasets only. These may be considered for ZAs working on the continuum land-sea following the same workflow.

4.2.2. Vegetation index

Abbreviated as NDVI, the normalized difference vegetation index differs from one season to another, depending on the plant growth and species (coniferous or deciduous trees for instance), the snow cover, the presence of urban fabric.

This greenness index is often used as a drought indicator since vegetation growth is driven by water during spring and summer. Inversely related to snow cover in winter, NDVI is also viewed as a climate change indicator in mountainous regions (Rumpf et al., 2022; Choler et al., 2021).

The MOD13C2 product provides monthly NDVI grids at the same resolution and the same frequency as the land surface temperature product. Processing the 237 NDVI grids (three grids are missing) is similar to processing the LST grids.

4.2.3. Population density

The Socioeconomic Data and Applications Center (SEDAC) provides world wide rasters (30 arc-second grid cells, WGS84) elaborated from national censuses. The SEDAC algorithm allocates population over sub national administrative units with a 5-year representation (2000, 2005, 2010, 2015, 2020). Such data may inform on the urbanization gradient.

The Population Density, v4.11 TIFF rasters produced by dividing the population number by the area are well documented. This information also come as CSV tabular files, assigning the population data to points. We consider 2000 and 2020 rasters.

Due to the resolution of 30 arc-seconds, computations are carried out on a personal computer (Intel® Core™ i7-9700 CPU @ 3.00GHz × 8 with 64 GiB of RAM memory).

The rasters are resampled at the resolution of 0.05° to be consistent with the MODIS sub-rasters. The rasters are cropped with respect to the ZA polygons. Mean densities (GPW4) in persons/km² are here averages computed at a resolution of 0.05°. A finer indicator could be computed before the change in resolution. A GPW4 indicator field is added to the ZAs' JSON file.

4.2.4. ESA WorldCover

Land cover products differ one from another in the classification and the resolution, hampering comparisons along time.

We select the recent ESA WorldCover product (Harper et al., 2022). This large dataset encodes 11 generic classes (trees, grassland, cropland, built-up, water bodies...) at a resolution of 10 m. Three macrotiles

(containing tiles of 3°×3°) covering the European continent and Zimbabwe are downloaded. As for population density, computation are carried out with 64 GiB of RAM memory.

Data are spread into ten rasters by aggregating the classes with respect to their ten. These rasters are then resampled. This allows to preserve class ratio information at a lower resolution. They are then cropped with respect to the ZA polygons. The class ratios (vector of size 10) are gathered under the LC indicator.

4.2.5. BioGeoRegions

The biogeographical regions dataset (shapefile) provides the delineations related to the Habitats Directive (92/43/EEC). More detail may be found in the related metadata sheet, the FAIRness of which is estimated in subsection 3.3.5.

The dataset is downloaded from the EEA geospatial data catalogue. As it comes with EPSG:3035 (Projected coordinate system for European Union countries and candidates), we use QGIS to re-project it to WGS 84.

Matlab is used to simplify the 12 polygons (notably along the sea coasts) and to save them in a unique raster using a tolerance of 0.05°. The raster is cropped with respect to the ZA polygons so as to determine their mix of habitats.

Metropolitan France is concerned with alpine (pink), Atlantic (blue), continental (green) and Mediterranean (orange) biogeographical regions, see Fig. 2. The ratios are added to the JSON files under the BIOG field.

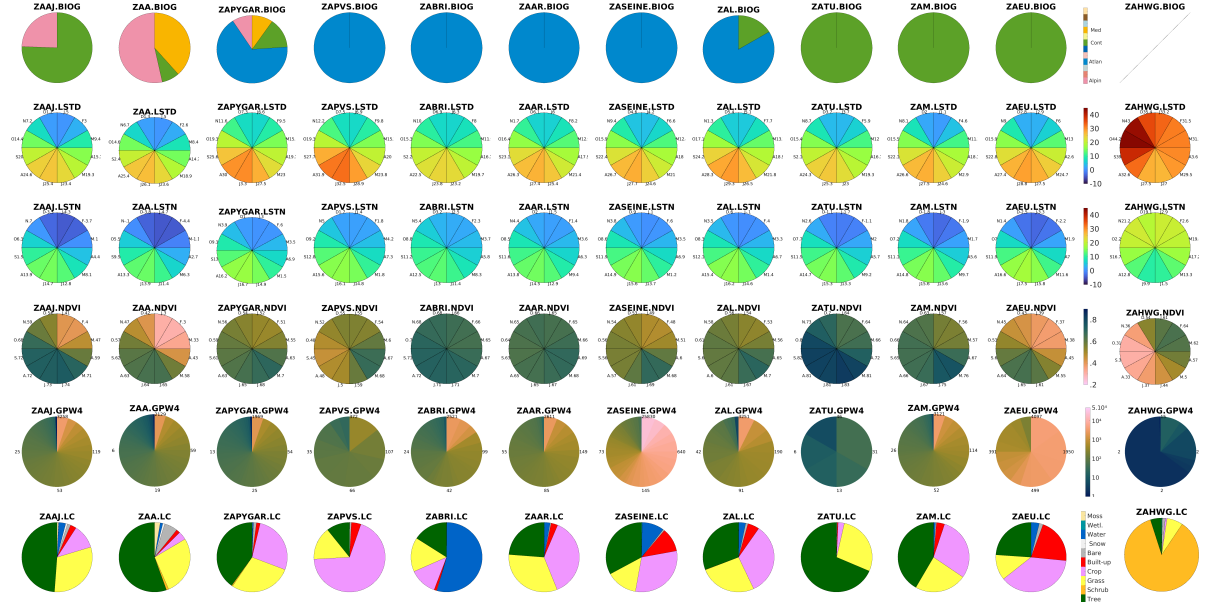
4.3. Indicator chart for an insight into gradients

As a reuse benchmark, indicators are read in the JSON files in order to plot a table of pie charts, Fig. 3, for 12 ZAs (columns) and 8 indicators (lines). This represents a total amount of 2449 values.

The ZAs are ordered clockwise following their biogeoregions (Alpine, Mediterranean, Atlantic, Continental), starting from the partly Alpine ZAs (ZAAJ and ZAA) located in the East of France, passing through South and West (partly) Atlantic ZAs (ZAPYGAR, ZAPVS, ZABRI, ZAAR, ZASEINE), then (partly) continental ZAs located in the center and the North East of France (ZATU, ZAM, ZAEU). Pies for ZAHWG (Zimbabwe, Africa) are provided too.

Indicator data are read in the ZAx.json files while color maps are read in the JSON files describing the variables. The color maps, namely turbo and batlow, are chosen to facilitate the visual comparison of the temperatures, vegetation and population pie charts. On the pies, a small font size is chosen to favour the visual impression against a focus on numbers.

Annual indicators (averages, quantiles, or ratios)



Climate change (temporal series)

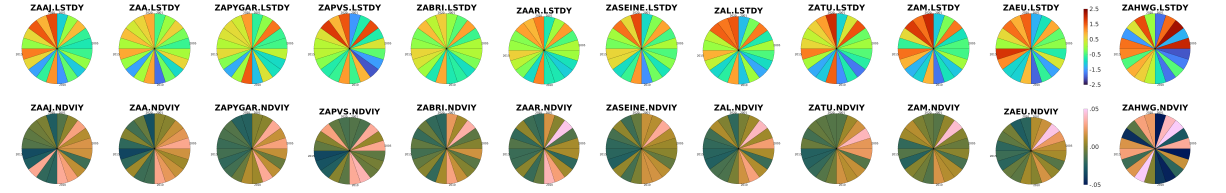


Figure 3: Indicators (lines) for 12 Zones Ateliers (columns). Lines: Biogeoregions ratios, LSTD month averages, LSTN month average, NDVI month average, GPW4 quantiles, Land cover class ratios, LSTD and NDVI variations over 2001-2020.

Biogeoregions (BIOG) and land cover (LC) pie slices represent ratios and are colored with respect to their classes. Note that comparisons between these two variables are somewhat hindered because of shared colors (notably blue) with different meaning. Six ZAs lie in a unique region, while the other six regions belong to two to four biogeoregions. One notices that DEIMS-SDR data report one biogeoregion per ZA only.

Monthly means for MODIS data (LSTD, LSTN, NDVI) allow to show the annual variation (2001-2020) of temperature and vegetation as 12-slices pie charts. A turbo color map, blue for cold months and red for hot months, is used for temperatures.

One observes that the Metropolitan ZAs present very similar land surface temperature pie charts, differing from one to another on some color tones that correspond to a few degrees. Located in Africa, ZAHWG is an outlier.

Mean temperatures computed over large areas with a high variation in the elevation (such as ZAA) under-

estimate or overestimate local temperatures. Note that DEIMS-SDR proposes a unique average temperature for the moment.

A flipped “batlow” colormap ranging from pink to green is used for the green NDVI index. These pies exhibit more variety. This can be explained by differences in climate, land cover and elevation, characteristics that are summarized into the biogeoregions.

As an example, in Brittany, ZAAR and ZABRI present very similar pie charts, at the exception of land cover. These qualitative and quantitative indicators could endorse the project of a common plate-form based on a land-sea continuum submitted by these neighboring ZAs.

In contrast, the fact that ZAA and ZAEU share similar LST and NDVI pie charts is misleading. The small urban ZAEU is greener in summer due to deciduous trees. These trees are leafless in winter and thus the urban fabric is visible. At the same time, some green areas in the large ZAA are covered by winter snow.

For the population density (GPW4), quantiles data are represented counterclockwise using the batlow colormap ranging from green (rural, sparse) to pink (urban, dense) and a logarithmic scale for a better rendering. One recognizes patterns for rural ZAs (ZAPVS and ZATU) and for large areas (ZAA, ZAAR, ZABRI, ZAAJ, ZAM, ZAPYGAR). The latter are mostly rural with several big cities concentrating people, historically set up on main rivers (see Fig. 2). This is confirmed by riverine ZAs (ZAL and ZASEINE) and the urban ZA (ZAEU) located on the Rhine river.

The values on the GPW4 pies indicate the average number of persons per km² for some of the slices. For instance, the density of inhabitants in the half of the ZAEU territory is less than 561 persons/km². The most populated slice of the territory, one raster cell in this case, hosts 4097 persons/km² in average.

The last two lines of Fig. 3 plot the annual means for LSTD and NDVI between 2001 and 2020 to illustrate the climate change. As expected, trends differ between the two decades. The second one is hotter and greener than the first one, and presents higher variations. Note that the data are already available in (Charpentier, 2023) for studies carried out along seasons.

4.4. Discussion

The general workflow of Fig. 1 aims at providing indicators computed on a common basis at the level of a social ecological research network such as the RZA. As it uses the DEIMS-SDR information (boundaries) provided by the site managers, it could be extended to the eLTER network. Thinking in biodiversity conservation (COP 15), networks such as UNESCO Biosphere Reserves or French “Conservatoires des Espaces Naturels” may also be interested.

The few variables considered here already provide relevant informations, even if the ZA areas are sometimes very large. In practice, most of the ZA’s studies are carried out in smaller territories. Given a ZA, the dataset of rasters deposited on InDoRES could be cropped with respect to smaller territories allowing for a fair capture of both the spatial heterogeneity and the ZA’s gradients.

For eLTER sites lying in a unique raster cell/pixel (roughly 5×5 km²), spatial averaging is equal to the cell values. In other words, provisioning the JSON file containing the indicators and the CSV data will be an effective and proportionate solution.

The research on a social-ecological system is rich of its on-site observations (social and ecological) and the co-working with local authorities, associations and citizens. Indicators deduced from rasters may be evaluated

against open ground-based datasets (Klein Tank et al. (2002) for instance) and on-site observations.

The concise, coarse but informative indicator chart may be compared to or supplemented by a principal component analysis on the data. This will be carried out later by considering indicators computed on smaller ZAs’ sites.

Data (we searched) are findable, accessible and information is available to reuse them. FAIRness is not fully satisfied since data are mostly collected without metadata.

The workflow and computer codes already implemented for a variety of variables and formats may be reused to deal with new raster sets. A light version called ModZA will be made available through GitHub for the reuse of the rasters and indicators.

5. Conclusion

The FAIRness of the data is a key concern for a scientific research. Although not FAIR within the scientific meaning, open world wide raster sets we consider are provided with sufficient information to be reused. To go one step beyond, the heterogeneity in their characteristics is smoothed by implementing a general workflow to rework them consistently (same resolution, same formats, metadata description...). Throughout the study, additional sites, variables and indicators were considered as cross validation tests for robustness.

The choice is made to manage the workflow implementation, the dataset construction, the dataset fairification and deposit, the design of the indicators, their representation and analysis and the writing of the paper at the same time as a guarantee of the consistency between the different research products. FAIRification is a valuable, but time consuming task.

Finally, a concise, coarse but informative indicator chart informs on the gradients over a social ecological research network, striving for research and open science opportunities.

Acknowledgment

The author thanks the Réseau des Zones Ateliers for the opportunity to be in charge of its Data and Open Science working group. During these last three years, fruitful discussions about FAIR management with the RZA’s “Bank the data together” team and the joint service BBEES (Bases de données Biodiversité, Ecologie, Environnements Sociétés) that manages InDoRES, this under the supervision of CNRS/INEE, have undoubtedly stimulated the writing of this scientific testimony

References

- 11