The seamless interoperability of geospatial freshwater tools

Sami Domisch[1], Merret Buurman[1], Thomas Tomiczek[1], Afroditi Grigoropoulou[1,2], Marlene Schürz[1,2], Jaime Garcia Marquez[1], Giuseppe Amatulli[3], Yusdiel Torres Cambas[1], Igor Ogashawara[1], Markus Konkol[4], and Vanessa Bremerich[1]

[1]Leibniz Institute of Freshwater Ecology and Inland Fisheries (IGB), Berlin, Germany
[2]Institute of Biology, Freie Universität Berlin, Germany
[3]Yale School of the Environment, Yale University, New Haven, CT, USA
[4]52°North Spatial Information Research GmbH, Münster, Germany

Date of submission: 23.06.2024

Abstract

The complex spatial structure and longitudinal connectivity of freshwater habitat mirrors the complicated geospatial processing of such data, where focusing on a given stream segment or lake requires expanding the spatial window of analysis also to hydrologically connected habitats. Yet, processing routines that address the connectivity within the freshwater continuum remain often ignored, resulting in inaccurate analysis results. After a phase of heavy data lifting and tool development given the Hydrography90m dataset, the GeoFRESH platform, the hydrographr R-package and the integration of lakes into the hydrography, we recognize that although "all data and tools are there", the interoperability that would allow users to easily integrate the data and tools in efficient workflows is lacking. We will address this challenge by proposing the following five tasks to streamline and interconnect our tool portfolio towards an enhanced user-experience, with an emphasis on a low-entry point: (i) an online point-and-click catchment delineation tool, (ii) an interactive online map editor for easy 'point-snapping', which is a largely underestimated yet critical source of potential errors, (iii) and an application programming interface (API), along with a set of functions that allow accessing GeoFRESH from R, i.e., within hydrographr, and other HTTP-supporting environments. Moreover, and given the recent EU Restoration Law targeting 25,000 km free-flowing rivers in Europe, we intend to (iv) integrate dams and barriers across Europe in the hydrographical network, and to offer interactive connectivity functionalities within both GeoFRESH and hydrographr for facilitating free-flowing river analyses also for non-programming users. Finally, (v) we will draft a suite of vignettes, tutorials and workflows that highlight the interoperability of the tools. The pilot project is in line with the larger NFDI4Earth future strategy in advancing and maturing services and tools for research communities. For the freshwater realm, this will be achieved by a seamless integration of standardized, high-resolution and scalable data and tools for analysis workflows anywhere in the world, and by addressing beginners and experienced users alike from biological and Earth System sciences.

I. Introduction

The spatial structure of freshwater ecosystems, covering both flowing and standing waterbodies and their connectivity, requires special workflows that require advanced geospatial processing skills which may not be necessarily part of the skill portfolio of freshwater-related Earth System scientists. Ignoring the connectivity however leads to an incorrect assumption of spatial independence among water

bodies, which in turn, cascades into oversimplified analyses regarding nutrient fluxes, sediment transportation, or species distribution and migratory patterns[1]. This underlines the need for easy-to-use tools and low entry-points for performing freshwater geospatial analyses.

The proposed pilot project builds upon our previous successful efforts in developing a new baseline in geospatial freshwater-specific data and tools. These include the Hydrography90m[2] dataset (visualization available at https://geo.igb-berlin.de/maps/351/view), the GeoFRESH online platform[3], the hydrographr R-package[4], and associated data soon to be published: the global lake catchments across the network (ongoing pilot project), and global high-resolution freshwater-specific environmental variables ("Environment90m", Garcia Marquez et al., in prep), which, all together, push the boundaries towards novel, high-resolution and scalable freshwater geospatial analyses.

These tools are however not yet truly interoperable which challenges their sustainable long-term structure, poses a major hurdle for analysis workflows and most importantly, limits the uptake by the research communities. Our vision is hence to extend the modular structure and add direct linkages and interfaces across our global data tools, which, based on the feedback by users we received, would represent the missing links for an improved uptake and would contribute the building blocks for facilitating international cooperation within the freshwater realm of NFDI4Earth. Our important preliminary tests regarding the feasibility of the planned modules allowed us to come up with the following five tasks described below for a one-year pilot project. Our pilot addresses mainly the "FAIR Big Data Pilot" track though it touches also on the principles of the other tracks.

I. Pilot description

(i) Point-and-click catchment delineation

Choosing a study area by simply clicking on the map and to obtain the upstream catchment within seconds represents a low-entry and important cornerstone for any geospatial freshwater analysis workflow. Users can choose from a suite of possible output options, ranging from a simple bounding box to a single upstream catchment polygon (Fig. 1). We expect that this functionality will be central for the GeoFRESH platform since it would – for the first time – allow users to pick their study area in the form of a catchment with one simple click, anywhere in the world. The technical implementation involves a PostgreSQL[5] database with the pgRouting[6] addon that enables the hydrological routing of the global sub-catchments and stream segments. For users who wish to create a scripting workflow, this functionality is provided as a HTTP application programming interface (API) which conforms to the Open geospatial consortium (OGC) API specification for processing services. Users will have the possibility to interact via HTTP without registration or login for a seamless workflow integration.

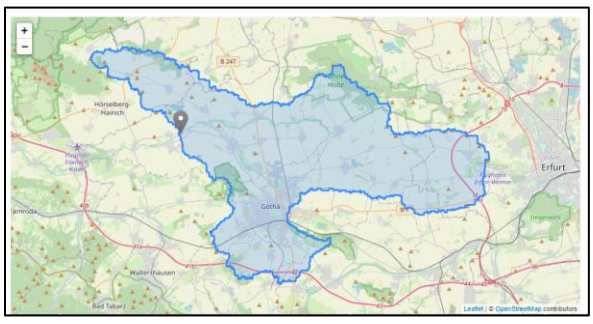

Figure 1 The demo-version of the interactive point-and-click catchment delineation tool, available at https://aqua.igb-berlin.de/upstream. By clicking on the map, the upstream-contributing area (in blue) can be visualized at any given location on Earth within seconds. This tool requires yet to mature and will be integrated into the GeoFRESH platform.

(ii) Interactive map editor in GeoFRESH

A central and daunting task in freshwater geospatial analyses involves integrating point data from e.g. water monitoring or biodiversity samplings to the stream network and lakes. The challenge lies in overcoming spatial mismatches between the point coordinates and the modelled stream network and lake shapes, where even a minor spatial offset may lead the point being attributed to an incorrect river segment or lake – an error consequently cascading to subsequent analyses. After receiving requests from the research community, we identified this key functionality as a major gap in the present GeoFRESH platform. Our proposed solution addresses, beside multiple automatic snapping possibilities, also the manual moving, editing and deletion of user-uploaded point data. The technical implementation will consist of an interactive frontend built with Leaflet and Shiny, communicating with the backend based on PostgreSQL. In addition, we will process more hydro-morphological background layers allowing custom visualizations to help users to identify important landmarks (e.g. given flow accumulation, direction, river confluences), and to allow exporting the map view. We believe that such functionality will significantly improve the user-friendliness and applicability of point-data processing while capitalizing on the online GeoFRESH infrastructure.

(iii) *API for connecting GeoFRESH and hydrographr*

While GeoGRESH is for non-programming users, we also offer the hydrographr R-package for users who wish to create custom workflows. Previous experience has shown that the data-intense and scalable processing, though feasible, can be considered a challenge by users whose personal computers do not offer the required capacity. We will address this issue by offering the possibility of a seamless interoperability such that users can tap on the GeoFRESH platform within R, but also other languages such as Python or Julia. The technical implementation requires a server platform where the processing is running, and which exposes an HTTP OGC API through which e.g. the R session on the user's computer triggers the processing and retrieves the results in a suitable format. The clear advantage lies in avoiding downloading large amounts of data input data – instead, the users will only download the processed result, or even leave the result on the server for further analyses. This is particularly useful when smaller data points have to be extracted from large input data, e.g. the connection between two points on the network, where only the connecting path needs to be returned.

(iv) *Integrating dams and barriers in GeoFRESH*

The EU Green Deal aims to restore free-flowing rivers in Europe which leads to the critical question of "Which rivers should be restored?". Answering this question requires first of all efficient data processing capabilities to assess European streams and rivers in concert. Our aim is to pro-actively build additional modules in GeoFRESH that take the burden of such data-intense processing routines from the research communities. Regarding the technical implementation, we propose (i) to integrate latest European dam and barrier data by tapping on the AMBER dataset[7], comprising 630,000 observed dams across Europe. In addition, (ii) we will capitalize on the powerful pgRouting extension of PostgreSQL and add additional point-connectivity calculation modules in GeoFRESH. We expect that such large-scale, pan-European distance-analyses between e.g. dams but also biodiversity features creates a valuable approximation regarding areas of high relevance for restoring free-flowing rivers.

(v) *New vignettes, tutorials and workflows*

Finally, the best tools will not reach their full potential if they are not conveyed to the research communities. Given the user feedback and requests, we will draft new example case study workflows and provide these within GeoFRESH and hydrographr, with a special emphasis on a low-entry point for users who are not familiar with geospatial analyses. This includes especially the automated and interactive point-snapping functionality and connectivity workflows. Workflows will be saved on the Aqua Galaxy platform and uploaded on Zenodo for an improved findability and reusability.

II. Relevance for the NFDI4Earth

Each of the five tasks outlined above represents a stand-alone implementation, and the strength lies in their combination which will significantly increase the potential uptake by research communities. The planned tasks mirror the actual need of researchers within the NFDI4Earth and NFDI4Biodiversity communities, underlining the high task relevance for the water-related research domains. We believe that pro-active preparations prior "the big run" on EU-wide connectivity analyses, to be expected within the next years, provides a promising avenue where NFDI4Earth will stand out as a critical entry point for international users. The will continue our efforts in bridging Earth System and biodiversity sciences, and we expect scientists, data curators, university teachers as well as the public authority to be stakeholders that can directly benefit from the pilot project. Moreover, the pilot project will intensify valuable synergies with the European Open Science Cloud (EOSC) and the ongoing AquaINFRA project (https://aquainfra.eu/) which creates virtual research environments tailored towards aquatic sciences. The pilot project tackles all elements of the FAIR criteria, as all data and code will be publicly stored in the IGB's Freshwater Research and Environmental Database, FRED, and on GitHub, respectively.

III. Deliverables

The main deliverable will be the completed tasks outlined above and hence (i) a seamless interoperability between the Hydrography90m-related data and tools, with workflows stored in Aqua Galaxy. In addition, we will (ii) produce a roadmap entitled "Seamless interoperability of large-scale freshwater geospatial processing tools" and contribute to the NFDI4Earth living handbook.

IV. Work Plan & Requested funding

The tasks of the one-year pilot project (Table 1) will be undertaken by a geospatial programmer. We request a budget of EUR 100.000 which covers the salary and a travel budget for outreach activities.

Work phases of the pilot project	Q1			Q2			Q3			Q4		
	1	2	3	4	5	6	7	8	9	10	11	12
Catchment delineation tool												
Interactive map editor												
API for hydrographr												
Barrier data & connectivity												
Tutorials / vignettes												
Roadmap document												

Table 1 Planned work phases of the pilot project for the programmer (shaded in blue). Q refers to quarters; numbers refer to months.

References: [**1**] Wohl, E. (2017) *Prog Phys Geogr* 41, 345-362 [**2**] Amatulli, G. *et al.* (2022) *Earth Syst. Sci. Data* 14, 4525-4550 [**3**] Domisch, S. *et al.* (2023) *Zenodo* [**4**] Schürz, M. *et al.* (2023) *Meth Ecol Evol* 14, 2953-2963 [**5**] PostgreSQL (2024) https://www.postgresql.org/ [**6**] pgRouting (2024) https://pgrouting.org/ [**7**] Belletti, B. *et al.* (2020). *Nature* 588, 436-441.