A Participatory Science Approach to Expanding Instream Infrastructure Inventories

Aaron Whittemore1, Matthew R. V. Ross2, Wayana Dolan3, Theodore Langhorst3, Xiao Yang3, Sayali Pawar4, Michiel Jorissen5, Eric Lawton6, Stephanie Januchowski7, and Tamlin Pavelsky3

1Geography, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA, 2Ecosystem Science and Sustainability, Colorado State University, Fort Collins, CO, USA, 3Geological Sciences, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA, 4Biosciences, Swansea University, Swansea, UK, 5Independent Research Aide, Eindhoven, The Netherlands, 6Independent Research Aide, Pontypool, Ontario, Canada

Abstract Over the past decade, remote sensing data have improved in resolution and become more widely available, bringing new opportunities for its use in environmental science and conservation. One potential application is to identify and map instream infrastructure across the world, with important implications for fisheries, hydrology, flooding, and more. To date, databases of instream infrastructure focus on larger dams with reservoirs that are comparatively easy to detect with remotely sensed imagery. Despite their impact on freshwater ecosystems, smaller infrastructure is often overlooked. To overcome these challenges, we require more systematic approaches, such as the Global River Obstruction Database (GROD) presented here, to map instream infrastructure. We present a participatory approach to identify, map, and validate infrastructure and provide an initial data set for the contiguous United States (n = 4,197). We highlight the value of participatory methods that include the public and suggest ways they could be fused with machine learning for future applications.

Plain Language Summary In recent years, imagery retrieved from Earth-observing satellites has improved in quality and become more widely available. We can use the improved satellite imagery to observe the Earth’s surface in entirely new ways. One potential application is to identify and map dams and other river obstructions that represent disturbances to the Earth’s freshwater systems. Keeping record of river obstructions is valuable for many purposes from fisheries management to understanding flood dynamics. Currently, global dam databases focus on larger structures, while smaller dams and infrastructure are often overlooked. Small river obstructions greatly outnumber their larger counterparts and collectively have a significant impact on freshwater ecosystems. To further knowledge on the location and type of global river obstructions, we created the Global River Obstruction Database (GROD). GROD uses a uniform approach that maps river obstructions across the globe regardless of size or reservoir presence. We present and validate our citizen science-based approach to map obstructions and provide an initial data set of 4,197 obstructions for the contiguous United States. We also highlight the value of incorporating members of the public in data collection and suggest ways citizen science-based methods could be fused with more technical applications like machine learning for future projects.

1. Introduction

Remote sensing of Earth’s surface has grown into an advanced field, providing high spatial resolution data across the globe (Campbell & Wynne, 2011). These remotely sensed data are increasingly available to diverse users via internet-based services (e.g., Google Earth), along with publicly available computational power to analyze large data sets (e.g., Google Earth Engine [GEE]; Gorelick et al., 2017). The combination of higher-resolution data and improved data accessibility is, in part, inspiring new approaches to data analysis; such methods are often referred to as either machine learning (ML) or artificial intelligence (AI). For example, ML and AI algorithms use vast stores of data to “learn” how to identify different objects, including human modifications to Earth’s surface, such as instream infrastructure like road culverts, bridges, or dams (e.g., Weil, 2018). Instream infrastructure are barriers to water flow that cause disturbance of aquatic ecosystems (Poff et al., 1997), and efforts to further catalog such structures are needed to help inform global river restoration and conservation initiatives (Neeson et al., 2015; Poff & Hart, 2002).
ML and AI approaches are most often used to identify relatively uniform features (e.g., Craciun & Zerubia, 2013; Tayara et al., 2017; Weil, 2018), and they require spatially distributed, high-quality training and validation data, often called “labeled” data. These data are rarely available for instream infrastructure, requiring large, unified efforts to build these labeled data sets. Simultaneously, researchers and members of the public work together through participatory (Miller-Rushing et al., 2012), or citizen, science to collect, aggregate, and process Earth observation data (Dickinson et al., 2012; Mulligan et al., 2020). Here, we refer to such collaborations as participatory science, with an emphasis on research or development projects, where researchers and members of the public work together, in near real time. These participatory projects are particularly useful for cataloging broad-scale inventories of human-built infrastructure like structures and roads identified in open street map (Goodchild, 2007) and show potential for training ML or AI methods (e.g., George, 2019).

Data generated through participatory science projects can be coupled with ML or AI methods to identify structures over broader areas, more quickly, and with greater reproducibility (e.g., George, 2019). However, despite some robust catalogs of certain types of infrastructure (viz., larger dams and roads), there are relatively few complete inventories of instream infrastructure (Januchowski-Hartley et al., 2013; Poff et al., 1997). Refining how we identify and inventory instream infrastructure is a global conservation priority (Grill et al., 2017). Mapping the spatial locations of different infrastructure types is a primary requirement for spatially explicit river restoration planning and to identify more cost-effective solutions (Neeson et al., 2015) that also return benefits to species of conservation interest (Hermoso et al., 2018; Milt et al., 2018).

The GRanD (Lehner et al., 2011) and AQUASTAT (Food and Agriculture Organization [FAO], 2016) databases offer global coverage, draw on various sources to catalog large (>15 m high) dams, and have varying levels of spatial accuracy. The approaches used to collate and create GRanD and AQUASTAT are challenged by the variety of methods and criteria used to map and classify instream infrastructure across their component, regional databases. Merging existing regional databases can cause over or under representation of infrastructure, with the latter particularly true for areas of the globe where no comprehensive monitoring effort has been undertaken.

Other existing global inventories were developed by bringing together ML and remotely sensed imagery to identify and map dams and associated reservoirs across the globe. Khandelwal et al. (2019) used ML and Landsat imagery to identify reservoirs associated with dams built between 1984 and 2012. Weil (2018) created a ML pipeline during the 2018 GeoForGood User Summit hackathon to map dams using Sentinel-2 10 m imagery. Both approaches used by Khandelwal et al. (2019) and Weil (2018) were limited to large dams and reservoirs and are not suited for locating smaller obstructions because of their dependency on reservoir presence or training data from existing dam inventories, which do not include smaller structures. Finally, the GOODD database (Mulligan et al., 2020) depended on members of the public and researchers coming together to digitize visible dams on satellite imagery in Google Earth with a minimum reservoir length of 500 m and dam wall length of 150 m. Although the GOODD method has challenges with regard to accuracy and consistency of participants mapping infrastructure, and spatial resolution was limited to that available on Google Earth at the time of mapping (2007–2011), we believe there is value in further developing this participatory approach to map smaller infrastructure. A global database of smaller infrastructure could serve as training data for methods like those used by Khandelwal et al. (2019) and Weil (2018), effectively easing previous limitations to cataloging small instream infrastructure via ML.

Here, we present methods, validation, and initial results for the Global River Obstruction Database (GROD), which is being generated through a participatory science project that uses remotely sensed imagery in GEE. Once completed, GROD will be global in scale and include instream infrastructure of different types and sizes that could act as obstructions to flows of water, material, or species, on rivers defined by the Global River Widths from Landsat (GRWL) database (Allen & Pavelsky, 2018). We present an initial validation of instream infrastructure mapped by GROD participants in two regions, one in the United States and one in France. We also present an initial GROD data set for the contiguous United States. Further, we explore the role of the completed data set in addressing global-scale questions about the effects of different infrastructure on hydrology and ecology. Finally, we discuss how GROD data and approach can be used to inform and validate future ML methods that expand and improve global mapping of instream infrastructure and how similar approaches combining participatory science and ML could be used more broadly to expand our knowledge of the Earth system.
2. Materials and Methods

2.1. GROD Approach

Mapping instream infrastructure for the GROD was initially undertaken by 13 people. These initial GROD participants were recruited through personal connections or by websites such as Twitter (via the hashtag #DamOrNot) and SciStarter. All GROD participants accessed GEE (Gorelick et al., 2017) via a web browser to map infrastructure.

At the start of a mapping session, each participant loaded GEE’s basemap satellite imagery, the GRWL river data, and any infrastructure already mapped by that user into GEE using bespoke code for the project. GRWL includes mapped centerlines for >2.1 × 10⁶ km of rivers wider than 30 m at mean annual discharge (Allen & Pavelsky, 2018). The bespoke code for the GROD project divides the GEE imagery into 1,039 tiles that are 6°E to 6°W and 3°N to 3°S. Each tile intersects with GRWL centerlines so that there are no empty tiles. Most GEE imagery has submeter spatial resolution, which was determined to be adequate for identifying the majority of instream infrastructure on GRWL rivers.

Within an individual tile, participants scrolled along GRWL river centerlines and, using the tools within GEE, mapped a point feature on top of any observed infrastructure (Figure S1 in the supporting information) along the river. Participants classified each mapped infrastructure into one of seven predefined classes, grouping by similar flow alterations and visual characteristics (Figure 1). Our goal with these classes was to strike a balance so that participants could distinguish among classes while infrastructure is grouped in a meaningful way based on likely flow alteration. The classes from maximum to minimum flow alteration were dams (infrastructure that covers the entire river channel, allowing no water to pass except through the structure itself), locks (a subset of dams that are distinguished by their lock mechanism that allows barges and other river traffic move up and downstream), channel dams (any obstruction on a multichannel section of river that does not obstruct all channels), partial dams ≥50% (infrastructure that is impermeable and covers more than half, but not all of the river channel), partial dams ≤50% (infrastructure that is impermeable and covers less than half of the river channel), low permeable dams (infrastructure that has a small height difference upstream and downstream of the structure [i.e., low head dams] or allows water to pass through), and uncertain (the type of infrastructure is not discernible). All classes of instream infrastructure were defined with examples provided in the tutorial document shared with participants before they commenced mapping. To facilitate the mapping and classification of infrastructure, we also asked all participants to review a decision tree (Figure 2) and a tutorial (see https://globalhydrologylab.github.io/GROD/).

At the end of every mapping session, participants exported all instream infrastructure mapped in GEE to a CSV file that included index, class, and latitude and longitude information for all mapped structures. Once a participant mapped all infrastructure on GRWL river lines in a cell, it was flagged as completed on GEE.

2.2. Validation

We validated instream infrastructure mapped by the 13 participants for four river reaches: two in France and two in the United States (Figure S2), drawing on existing regional-scale inventories of infrastructure for the two areas as well as a participant-generated truth data set (PGTD). These four river reaches were selected because (1) an expert user mapped instream infrastructure in the areas, (2) the areas were known to have high infrastructure density and variety, (3) data were available from regional-scale instream infrastructure inventories (detailed below), and (4) completion times were short enough (~1 hr) to allow for repeatability and comparison. For the validation procedure, participants identified and classified infrastructure over the same four river reaches for comparison purposes. All participants mapped all infrastructure on GRWL river lines in a cell, it was flagged as completed on GEE.

We validated instream infrastructure mapped by the 13 participants for four river reaches: two in France and two in the United States (Figure S2), drawing on existing regional-scale inventories of infrastructure for the two areas as well as a participant-generated truth data set (PGTD). These four river reaches were selected because (1) an expert user mapped instream infrastructure in the areas, (2) the areas were known to have high infrastructure density and variety, (3) data were available from regional-scale instream infrastructure inventories (detailed below), and (4) completion times were short enough (~1 hr) to allow for repeatability and comparison. For the validation procedure, participants identified and classified infrastructure over the same four river reaches for comparison purposes. All participants mapped all infrastructure on GRWL river lines in a cell, it was flagged as completed on GEE.

The two regional inventories included in our validation analyses were the Southeast Aquatic Connectivity Assessment Project (SEACAP; Martin et al., 2014) for the two river reaches in United States and the National Repository of Obstacles to Flow (ROE) for the two river reaches in France. We drew on a modified version of the publicly available ROE database (see Januchowski-Hartley et al., 2019) that excluded destroyed, planned, under construction, invalid, or duplicate infrastructure. We created the PGTD data set by creating manual buffers around each unique instream infrastructure location and then removed features in the lowest quartile of interparticipant accuracy, which primarily removed infrastructure classified as “uncertain.” This approach resulted in 93 PGTD-derived instream infrastructure locations for the
validation (down from an original 149 locations). Of the original 149 unique instream infrastructure locations mapped by GROD participants on GRWL river reaches in the two areas used for our validation, only 58 overlapped within manual buffers of obstructions recorded in SEACAP (7 obstructions) or ROE (51 obstructions). Of these 58 instream infrastructure, 3 did not overlap with the PGTD (i.e., they were in the lower quartile of the original 149 instream infrastructure found by GROD participants), resulting in 96 distinct instream infrastructure locations used for comparison in our hit rate analysis.

Figure 1. Satellite imagery examples of each unique class featured in GROD (uncertain excluded due to variability). All imagery provided by GEE.
We used the two regional inventories and the PGTD to calculate a “hit rate” or the percent of GROD participants who correctly mapped infrastructure at each location that occurred in the validation data sets. Drawing on all 149 instream infrastructure mapped by the 13 GROD users, we used a confusion matrix analysis to determine overall accuracy (for all mapped locations), as well as sensitivity, specificity, and precision for each of the seven infrastructure types (e.g., dam and lock) in the two validation areas (France and the United States). User predictions can take one of four classifications in a confusion matrix output: a true positive (TP) in which a user predicted an infrastructure class at a location which matched the actual class, a true negative (TN) in which a user predicted that an infrastructure class did not exist at a location and the actual class agreed, a false positive (FP) in which a user predicted class at a location was different than the actual class, and a false negative (FN) in which a user did not predict an infrastructure type at a location but the actual value did. The four metrics were calculated as follows: overall accuracy = (TP + TN)/(TP + TN + FP + FN); sensitivity = TP/(TP + FN); specificity = TN/(TN + FP), and precision = TP/(TP + FP).

Figure 2. Decision tree provided to all GROD participants for training and decision making related to classification of an identified instream infrastructure. All imagery sourced from Google Earth Engine.
The validation procedure served as an indicator of our method's performance for identifying and classifying infrastructure and helped refine the method for subsequent iterations of mapping.

### 2.3. GROD Data for U.S. Rivers and Comparison With Existing Data Sets

GROD participants identified, mapped, and classified instream infrastructure on GRWL rivers in the contiguous United States. Participants were trained and familiar with the GROD mapping process prior to contributing to contiguous U.S. data collection. Unlike the validation data set described above, there was no duplicate mapping of infrastructure in this process. Instead, each participant mapped obstructions in a tile and, once completed, moved on to another unfinished tile to continue mapping. In addition to the validation procedure explained above, we also compared the instream infrastructure mapped by GROD participants for GRWL rivers in the contiguous United States with those in the GRanD, GOIDD, and SEACAP inventories.

## 3. Results

### 3.1. GROD Validation

Our validation showed that GROD participants had perfect identification of instream infrastructure (100% hit rate) when compared with those mapped by SEACAP (n = 7) and a 92% (mean) hit rate when compared with ROE (n = 51) (Table 1). We also found that regardless of the measure (mean or median), combined results from the United States and France had >80% hit rates (Table 1). Comparison of data from GROD participants with the PGTD (n = 36, United States; n = 57, France) produced a lower mean (70%) and median (78%) hit rate than with SEACAP and ROE (82% and 92%, respectively; Table 1).

Overall users agreed on the obstruction class 70% of the time based on a confusion matrix analysis. Looking more specifically at individual infrastructure classes, we found that dams, channel dams, and low permeable dams had lower (<70%) precision ratios. Notably, we found that dams and channel dams had the lowest sensitivity (58% and 39% respectively; Table 2), and low permeable dams had relatively low specificity (82%; Table 2). We also found that participants frequently classified dams as low permeable dams and vice versa (Table S1), highlighting difficulties of assessing dam height from 2-D imagery.

### 3.2. Instream Infrastructure in the United States

Five participants mapped 4,197 instream infrastructure features on 108,993 km of GRWL rivers across 67 tiles in the contiguous United States using the GROD approach (Table 3). We found that GROD participants mapped 15 times more instream infrastructure than documented in GRanD, 13 times more than GOIDD (Figure 3), and 3 times more than SEACAP (Figure 4). Of the infrastructure mapped by GROD participants,

| Class               | Sensitivity | Specificity | Precision | Mean # identified by user | # of PGTD locations |
|---------------------|-------------|-------------|-----------|---------------------------|---------------------|
| Channel dams        | 39%         | 94%         | 44%       | 8                         | 14                  |
| Dams                | 58%         | 92%         | 58%       | 13                        | 14                  |
| Locks               | 77%         | 93%         | 77%       | 17                        | 26                  |
| Low permeable dams  | 75%         | 82%         | 67%       | 26                        | 32                  |
| Partial dams >50%   | 84%         | 100%        | 90%       | 4                         | 6                   |
| Partial dams <50%   | 22%         | 100%        | 100%      | 1                         | 2                   |
| Uncertain           | 79%         | 99%         | 90%       | 12                        | 55                  |
| Total               |             |             |           |                           | 149                 |

| Data set                        | United States | FRANCE | Overall |
|---------------------------------|---------------|--------|---------|
| # of instream infrastructures   | SEACAP   | PGTD   | ROE     | PGTD   | SEACAP/ROE | PGTD   |
| Median hit rate                 | 100%      | 33%    | 92%     | 92%    | 92%        | 78%    |
| Mean hit rate                   | 100%      | 51%    | 79%     | 82%    | 82%        | 70%    |
we found that 811 were classified as dams and that these most closely represented those contained in GRanD (Lehner et al., 2011). GRanD contained 206 of these 811 dams, which accounted for 84% of total overlapping infrastructure between the data sets.

4. Discussion

4.1. Strengths and Limitations of the GROD Approach

Our results suggest that GROD provides more detail than existing databases of instream infrastructure over the contiguous United States, whether they are global databases like GRanD and GOodd or more regional databases like SEACAP. Our findings highlight the need for a more complete global database of instream infrastructure, even in regions with relatively good records such as the contiguous United States. Given the long history of human modification of riverscapes, the presence of unrecorded infrastructure that may no longer be actively managed is perhaps unsurprising. An approach such as that presented here can help to address this deficiency.

Still, there are limitations to the GROD approach. Lower hit rates in the PGTD comparison, and in the United States compared to France, were largely driven by a higher number of less obvious infrastructure classes being identified, such as those classified as “uncertain” or “partial dams” by GROD participants (Table S2). Structures in SEACAP and ROE were easier to identify, as they generally include larger infrastructure such as locks and dams with reservoirs that are easier to visually discern. Findings from the PGTD comparison illustrate the difficulties associated with locating and discerning smaller infrastructure that are often not included in existing data sets.

Channel dams had low sensitivity, which can be partially explained by the difficulty for participants to identify if the area in question is one channel of a multichannel river or not. Additionally, sometimes dams are situated at the start or end of a channel or branch of the mainstem of the river, and this can make it difficult to determine if the structure should be classified as a “channel dam” or another class. The introduction of

| Class               | Count | Percentage |
|---------------------|-------|------------|
| Channel dams        | 383   | 9%         |
| Dams                | 811   | 19%        |
| Locks               | 238   | 6%         |
| Low permeable dams  | 436   | 10%        |
| Partial dams <50%   | 1,436 | 34%        |
| Partial dams >50%   | 394   | 9%         |
| Uncertain           | 499   | 12%        |

Table 3
Summary Statistics of Different Infrastructure Class Types Mapped for Rivers in the United States

Figure 1. GROD data compared to GRanD and GOodd for GRWL rivers in the contiguous United States. Darker circles indicate general increase in obstruction of flow. GRanD and GOodd dams are identified as light gray triangles and dark gray diamonds, respectively.
multiple steps into classifying channel dams could have been another factor influencing inconsistency between users. Participants also struggled to differentiate between dams and low permeable dams, which underscores differences in visual interpretation of imagery or the GROD decision tree among participants.

The inconsistencies noted above are limitations of the participatory approach used to generate GROD, and of participatory approaches more broadly (Theobald et al., 2015). Although the imagery was of high quality for the validation regions in France and United States, the high variety and density of infrastructure along validation reaches likely present a near worst-case scenario in terms of interparticipant consistency. Despite these challenges, overall classification agreement remained relatively high in the validation regions (70%).

The initial validation was limited by a relatively small number ($n = 13$) of participants. Completion times varied, but participants usually finished in about 1 hr. Performing validation on shorter river sections would have reduced completion time and perhaps attracted more validation participants, but having each individual participant classify a diverse range of structures was informative, even if requiring longer river sections and completion times.

While the number of participants was low, they were diverse in background and age, and many had never been exposed to the GROD application before participating in the validation effort. Even with variable prior experience, all participants completed the validation and showed reasonable consistency in classification and identification of instream infrastructure. From our experience, the validation results instilled confidence that participants with limited prior experience can contribute to collecting GROD data in future iterations of mapping. Finally, our project is still ongoing, and the initial validation steps will allow us to continue to improve our training and decision tree for subsequent mapping and project iterations.

As the GROD project grows, we anticipate additional participants and the potential for further differences in interpretation of the project guidelines and GEE imagery. At the same time, the GROD approach is focused...
on identifying instream infrastructure that is not typically included in other global-scale databases. The overall accuracy of participants in mapping different classes of infrastructure suggests that the GROD database will be useful for expanding on existing inventories. As a result, we decided to maintain a variety of different classes, leaving it to end users to group classes that remain difficult to differentiate during post hoc evaluation as needed.

Given the high variety of infrastructure and natural ambiguity in class attribution, some disagreement was expected among participants, but through our validation checks, we have demonstrated that participants can effectively identify infrastructure and agree on their classification in the majority of instances. Our approach to mapping infrastructure is well suited for broad participation and could lead to faster completion of infrastructure identification for the entire globe, along with crowd-sourced verification and validation. In result, GROD could be an ever-updating database of instream infrastructure as new imagery is released. Critically, these updates will be tracked with version control information that is clear and readily available in figshare, our chosen data repository. Anytime GROD is updated, improved, or altered, these changes will be documented in the same figshare link below, but with a new version number highlighting the alterations. This is common practice for dealing with changes and improvements in training data as outlined by the research data alliance (https://www.rd-alliance.org/data-versioning-wg-final-recommendations-and-next-steps).

4.2. Looking to the Future: Fusion of Participatory and ML Approaches

To both expand and improve GROD and similar data sets, we see great potential in merging participatory approaches, which can be accurate but rather time intensive, with ML and AI approaches that are fast but require large training and testing data sets. Training and validation data sets are vital in other fields of image recognition applications, such as facial recognition (Phillips et al., 2005). Most facial recognition software relies on so-called “labeled” data or images where faces and their expressions have been identified by people (Huang et al., 2008). The “labeled” data are used to train ML algorithms that can then encounter an unlabeled image and find the faces in it.

In a similar way, GROD participants have mapped and classified thousands of instream infrastructures. Because it includes a much broader array of riverine obstructions, GROD will be more useful than existing global-scale data sets and more likely to support ML approaches that will succeed in identifying larger (dams) and smaller (partial dams, low permeable dams, or locks) structures (sensu Sheng et al., 2008). Because our classifications are already defined by optical differences, GROD should be well suited to act as training data for ML (e.g., Craciun & Zerubia, 2013; Moranduzzo & Melgani, 2013; Tayara et al., 2017). Participants could validate predictions made by algorithms informed by GROD data and assist in areas of the globe where ML performs poorly. In addition, PLANET data (3 m resolution imagery) have been used to identify all roads and buildings in the world (George, 2019), based on OpenStreetMap training data, demonstrating that these finer-resolution imagery could assist with expanding the inventory of infrastructure captured by GROD, especially for smaller rivers. Finally, we see the potential for ML classification schemes to be applied to infrastructure already identified by GROD participants to create more consistent groupings and classifications.

We support participatory and open approaches to research and see broad possibilities for our approach and code (https://github.com/GlobalHydrologyLab/GROD) to inform and be a part of future fusions between participatory and ML methods. By developing an app-like front end to GEE, we made it possible for anyone in the world (including 13 participants from four different nations initially) to access terabytes worth of imagery, shapefiles, and other information. These data sets, combined, made it relatively straightforward for participants to scroll along rivers and identify infrastructure along the way. While our application focused on rivers, our code is easily modified for participants to hand-draw outlines of disturbances like beetle kill (Edenburg et al., 2012), fire (Jones et al., 2016), or hurricane wind fall (Baumann et al., 2014) and to use these outlines to train an algorithm that would expand the scope, accuracy, and speed for identifying such disturbances globally. Finally, our application could be modified to examine any environmentally relevant parameter of interest, and we note the potential to gamify this whole experience as a way to attract hundreds to thousands of more participants, like that presented in EyeWire (https://eyewire.org/explore) (Tinati et al., 2017).
5. Conclusion

At the global scale, our existing understanding of instream infrastructure impacts on sociohydrology and ecology are skewed toward those caused by larger infrastructure such as dams (see Grill et al., 2017; Liermann et al., 2020), but we increasingly understand that smaller infrastructure are more numerous and likely to have greater cumulative changes on our river ecosystems that remain poorly understood (Couto & Olden, 2018; Csiki & Rhoads, 2010; Januchowski-Hartley et al., 2013; Lange et al., 2019). By mapping infrastructure that does not create large reservoirs behind it, spatially explicit data sets like GROD will enable both finer-resolution analyses and better understanding of impacts caused by instream infrastructure on hydrological flow alterations (Poff & Hart, 2002), geographic range connectivity for species, particularly fishes (sensu Januchowski et al., 2020), and important fisheries resources (sensu Carvajal-Quintero et al., 2017). Equally, GROD will enable spatially explicit scenarios and potential solutions for restoring hydrological and ecological connectivity through the remediation (fish passages and structure modifications) or removal of existing infrastructure along global river networks (sensu Hermoso et al., 2018, Neeson et al., 2015). To date, GROD includes >4,000 classified instream infrastructure along GRWL rivers in the contiguous United States and expands upon existing global-scale inventories of infrastructure. We view the approach and data generated through GROD as initial steps toward fostering the fusion of participatory and ML approaches. Such joint approaches can turn big remote sensing data sets from raw potential into global information about instream infrastructure and assist us with quantifying and monitoring how our environments are changing over time. Once completed for the entire globe, GROD will be useful in coupling existing knowledge about river systems with new global surface water hydrology data sets derived from satellites, such as measurements of water surface elevation, inundation extent, and river discharge from the upcoming Surface Water and Ocean Topography satellite mission (Biancamaria et al., 2016).

Data Availability Statement

The original ROE data are available online (from http://www.onema.fr/le-roe), and the modified version is available from Januchowski-Hartley et al. (2019; https://figshare.com/s/617347a78cc27f419023). Both GRaND and GOODD data were obtained from GDW (http://globaldamwatch.org/grand/). River centerlines from GRWL are available from Zenodo (https://zenodo.org/record/1297443#.XmJn9ChKjZs). The GROD approach and code used to create the Google Earth Engine application for locating and identifying instream infrastructure are available from GitHub (https://github.com/GlobalHydrologyLab/GROD; DOI: 10.5281/zenodo.3722442). Finally, GROD validation and contiguous U.S. instream infrastructure data resulting from this project are available from figshare (https://figshare.com/articles/GROD_US_and_validation_data/12003513).

References

Allen, G. H., & Pavelsky, T. M. (2018). Global extent of rivers and streams. Science, 361(6402), 585–588.

Barbarossa, V., Schmitt, R. J., Huijbregts, M. A., Zarfl, C., King, H. M., & Schipper, A. M. (2020). Impacts of current and future large dams on the geographic range connectivity of freshwater fish worldwide. Proceedings of the National Academy of Sciences, 117(7), 3,648–3,655.

Baumann, M., Ozdogan, M., Wolter, P. T., Krylov, A., Vladimirova, N., & Radeloff, V. C. (2014). Landsat remote sensing of forest windfall disturbance. Remote sensing of environment, 143, 171–179.

Biancamaria, S., Lettemmaier, D. F., & Pavelsky, T. M. (2016). The SWOT Mission and Its Capabilities for Land Hydrology. Surveys in Geophysics, 37(2), 307–337. https://doi.org/10.1007/s10712-015-9346-y

Campbell, J. B., & Wynne, R. H. (2011). Introduction to remote sensing. New York, NY: Guilford Press.

Carvajal-Quintero, J. D., Januchowski-Hartley, S. R., Maldonado-Ocampo, J. A., Jézéquel, C., Delgado, J., & Tesedo, P. A. (2017). Damming fragments species’ ranges and heightens extinction risk. Conservation Letters, 10(6), 708–716.

Couto, T. B. A., & Olden, J. D. (2018). Global proliferation of small hydropower plants—Science and policy. Frontiers in Ecology & the Environment, 16(2), 91–100. https://doi.org/10.1002/fee.1746

Cracium, P., & Zerubia, J. (2013). Unsupervised marked point process model for boat extraction in harbors from high resolution optical remotely sensed images. Paper presented at the 2013 IEEE International Conference on Image Processing.

Csiki, S., & Rhoads, B. L. (2010). Hydraulic and geomorphological effects of run-of-river dams. Progress in Physical Geography, 34(6), 755–780.

Dickinson, J. L., Shirk, J., Bonter, D., Bonney, R., Crain, R. L., Martin, J., et al. (2012). The current state of citizen science as a tool for ecological research and public engagement. Frontiers in Ecology and the Environment, 10(6), 291–297. https://doi.org/10.1890/110236

Edberg, S. L., Hicke, J. A., Brooks, P. D., Pendall, E. G., Ewers, B. E., Norton, U., et al. (2012). Cascading impacts of bark beetle–caused tree mortality on coupled biogeochemical and biogeochemical processes. Frontiers in Ecology and the Environment, 10(8), 416–424. https://doi.org/10.1890/110173

Food and Agriculture Organization (FAO). (2016). AQUASTAT Georeferenced Dam Database. Food and Agriculture Organization of the United Nations.
George, M. (2019). Mapping all of Earth’s roads and buildings from space. https://www.planet.com/pulse/mapping-all-of-earths-roads-and-buildings-from-space/. Viewed 4 Nov 2019.

Goodchild, M. F. (2007). Citizens as sensors: The world of volunteered geography. GeoJournal, 69(4), 211–221.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote sensing of environment, 202, 18–27.

Grill, G., Lehner, B., Lumsdon, A. E., MacDonald, G. K., Zarfl, C., & Liermann, C. R. (2017). Reply to comment on ‘An index-based framework for assessing patterns and trends in river fragmentation and flow regulation by global dams at multiple scales’. Environmental Research Letters, 12(3), 038002.

Hernando, V., Filipe, A. F., Segurado, P., & Beja, P. (2018). Dealing with barriers in a systematic planning framework. Aquatic Conservation: Marine and Freshwater Ecosystems, 28, 17–25.

Huang, G. B., Mattar, M., Berg, T., & Learned-Miller, E. (2008). Labeled faces in the wild: A database forstudies face recognition in unconstrained environments.

Januchowski-Hartley, S. R., Jézéquel, C., & Tedesco, P. A. (2019). Modelling built infrastructure heights to evaluate common assumptions in aquatic conservation. Journal of environmental management, 232, 131–137.

Januchowski-Hartley, S. R., McIntyre, P. B., Diebel, M., Doran, P. J., Infante, D. M., Joseph, C., & Allan, J. D. (2013). Restoring aquatic ecosystem connectivity requires expanding inventories of both dams and road crossings. Frontiers in Ecology and the Environment, 11(4), 211–217.

Jones, G. M., Gutiérrez, R. J., Tempel, D. J., Whitmore, S. A., Berigan, W. J., & Peery, M. Z. (2016). Megadams: An emerging threat to old-forest species. Frontiers in Ecology & the Environment, 14(6), 300.

Khandelwal, A., Karpatne, A., Wei, Z., Kumbier, M., & McIntrye, P. B. (2011). High-resolution mapping of the world’s reservoirs and dams for sustainable river-flow management. Frontiers in Ecology and the Environment, 9(9), 494–502.

Khandelwal, A., Karpatne, A., Wei, Z., Kuang, H., Ghosh, R., Dugar, H., ... Kumar, V. (2019). GLADD-R: A new Global Lake Dynamics Database for Reservoirs created using machine learning and satellite data.

Lange, K., Wehrli, B., Abegg, U., Bätz, N., Brodersen, J., Fischer, M., ... et al. (2019). Small hydropower goes unchecked. Frontiers in Ecology and the Environment, 17(5), 256–258. https://doi.org/10.1002/fee.2049

Lehner, B., Liermann, C. R., Revenga, C., Vörösmarty, C., Fekete, B., Crouzet, P., ... et al. (2011). High-resolution mapping of the world’s reservoirs and dams for sustainable river-flow management. Frontiers in Ecology and the Environment, 9(9), 494–502.

Liermann, C. R., Nilsson, C., Robertson, J., & Ng, R. Y. (2012). Implications of dam obstruction for global freshwater fish diversity. BioScience, 62(6), 539–548.

Martin, E., Hoenke, K., Granstaff, E., Barnett, A., Kauffman, J., Robinson, S., & Apte, C. (2014). SEACAP: Southeast Aquatic Connectivity Assessment Project: Assessing the ecological impact of dams on Southeastern rivers. In The Nature Conservancy, Eastern Division Conservation Science, Southeast Aquatic Resources Partnership.

Miller-Rushing, A., Primack, R., & Bonney, R. (2012). The history of public participation in ecological research. Frontiers in Ecology and the Environment, 10(6), 285–290.

Mill, A. W., Diebel, M. W., Doran, P. J., Ferris, M. C., Herbert, M., Khoury, M. L., ... et al. (2018). Minimizing opportunity costs to aquatic connectivity restoration while controlling an invasive species. Conservation Biology, 32, 894–904.

Moranduzzo, T., & Melgani, F. (2013). Automatic car counting method for unmanned aerial vehicle images. IEEE Transactions on Geoscience and Remote Sensing, 52(3), 1,635–1,647.

Mulligan, M., van Soesbergen, A., & Sænsen, L. (2020). GOOaD, a global dataset of more than 38,000 georeferenced dams. Scientific Data, 7(1), 31. https://doi.org/10.1038/s41597-020-0362-5

Nesson, T. M., Ferris, M. C., Diebel, M. W., Doran, P. J., O’Hanley, J. R., & McIntyre, P. B. (2015). Enhancing ecosystem restoration efficiency through spatial and temporal coordination. Proceedings of the National Academy of Sciences, 112(19), 6,236–6,241.

Phillips, P. J., Flynn, P. J., Scruge, T., Bowyer, K. W., Chang, J., Hoffman, K., ... Worek, W. (2005). Overview of the face recognition grand challenge. Paper presented at the 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR’05).

Poff, N., Allan, J., Bain, M., Karr, J., Prestegaard, K., Richter, B., ... et al. (1997). The natural flow regime: A paradigm for conservation of riverine ecosystems. BioScience, 47, 769–784.

Poff, N. L., & Hart, D. D. (2002). How dams vary and why it matters for the emerging science of dam removal: An ecological classification of dams is needed to characterize how the tremendous variation in the size, operational mode, age, and number of dams in a river basin influences the potential for restoring regulated rivers via dam removal. BioScience, 52(8), 659–668.

Sheng, V. S., Provost, F., & Ipeirotis, P. G. (2008). Get another label?: Improving data quality and data mining using multiple, noisy labels. Paper presented at the Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining.

Svoboda, M., Haid, K. T. (2017). Vehicle detection and counting in high-resolution aerial images using convolutional regression neural network. IEEE Access, 6, 2,220–2,230.

Theobald, E. J., Ettinger, A. K., Burgess, H. K., DeBey, L. B., Schmidt, N. R., Froehlich, H. E., ... et al. (2015). Global change and local solutions: Tapping the unrealized potential of citizen science for biodiversity conservation. Biological Conservation, 181, 236–244.

Tinati, R., Luzak-Roesch, M., Simperl, E., & Hall, W. (2017). An investigation of player motivations in Eyewire, a gamified citizen science project. Computers in Human Behavior, 73, 527–540.

Well, C. (2018). Can we locate dams from space? https://medium.com/@charlotteweil/can-we-locate-dams-from-space-2a796ac8c04b. Viewed 28 Jul 2019.