ENVIRONMENTAL RESEARCH
LETTERS

PURPOSE-LED PUBLISHING™

Part of the ENVIRONMENTAL RESEARCH SERIES

TOPICAL REVIEW • OPEN ACCESS

A review of open data for studying global groundwater in social–ecological systems

To cite this article: Xander Huggins *et al* 2025 *Environ. Res. Lett.* **20** 093002

IOP Publishing Environ. Res. Lett. 20 (2025) 093002 https://doi.org/10.1088/1748-9326/adf127

ENVIRONMENTAL RESEARCH
LETTERS

RECEIVED
30 January 2025

REVISED
22 April 2025

ACCEPTED FOR PUBLICATION
17 July 2025

PUBLISHED
5 August 2025

TOPICAL REVIEW

A review of open data for studying global groundwater in social–ecological systems

Xander Huggins[1,2,3,4,*], Tom Gleeson[4,5], James S Famiglietti[6], Robert Reinecke[7], Daniel Zamrsky[8], Thorsten Wagener[9], Richard G Taylor[10], Megan Konar[11], Claudia Ruz Vargas[12], Miina Porkka[13], Lan Wang-Erlandsson[3], Inge de Graaf[14], Mark Cuthbert[15], Sara Lindersson[16], Yoshihide Wada[17], Marc F P Bierkens[8,18], Yadu Pokhrel[19], Juan Rocha[3], Giuliano Di Baldassarre[16], Matti Kummu[20], Grant Ferguson[21], Abhijit Mukherjee[22], Min-Hui Lo[23], Bridget R Scanlon[24], Mark S Johnson[1,25] and Chunmiao Zheng[26]

1 Institute for Resources, Environment, and Sustainability, University of British Columbia, Vancouver, BC, Canada
2 High Meadows Environmental Institute, Princeton University, Princeton, NJ, United States of America
3 Stockholm Resilience Centre, Stockholm University, Stockholm, Sweden
4 Department of Civil Engineering, University of Victoria, Victoria, BC, Canada
5 School of Earth and Ocean Sciences, University of Victoria, Victoria, BC, Canada
6 School of Sustainability, Arizona State University, Tempe, AZ, United States of America
7 Institute of Geography, Johannes Gutenberg-University Mainz, Mainz, Germany
8 Department of Physical Geography, Utrecht University, Utrecht, The Netherlands
9 Institute of Environmental Science and Geography, University of Potsdam, Potsdam, Germany
10 Department of Geography, University College London, London, United Kingdom
11 The Grainger College of Engineering, Civil and Environmental Engineering, University of Illinois at Urbana-Champaign, Champaign, IL, United States of America
12 International Groundwater Resources Assessment Centre (IGRAC), Delft, The Netherlands
13 Department of Environmental and Biological Sciences, University of Eastern Finland, Joensuu, Finland
14 Earth Systems and Global Change Group. Wageningen University and Research, Wageningen, The Netherlands
15 School of Earth and Environmental Sciences, Cardiff University, Cardiff, United Kingdom
16 Department of Earth Sciences, Uppsala University, Uppsala, Sweden
17 Biological and Environmental Science and Engineering Division, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia
18 Deltares, Unit Subsurface and Groundwater Resources, Utrecht, The Netherlands
19 Department of Civil and Environmental Engineering, Michigan State University, East Lansing, MI, United States of America
20 Water and Development Research Group, Aalto University, Espoo, Finland
21 Department of Civil, Geological, and Environmental Engineering, University of Saskatchewan, Saskatoon, SK, Canada
22 Department of Geology and Geophysics and School of Environmental Science and Engineering, Indian Institute of Technology Kharagpur, Kharagpur, India
23 Department of Atmospheric Sciences, National Taiwan University, Taipei, Taiwan
24 Bureau of Economic Geology, Jackson School of Geosciences, University of Texas at Austin, Austin, TX, United States of America
25 Department of Earth, Ocean and Atmospheric Sciences, University of British Columbia, Vancouver, BC, Canada
26 Eastern Institute for Advanced Study, Eastern Institute of Technology, Ningbo, People's Republic of China
* Author to whom any correspondence should be addressed.

E-mail: xander.huggins@ubc.ca

Keywords: global groundwater, open data, social–ecological systems, Earth systems

Abstract

Global data have served an integral role in characterizing large-scale groundwater systems, identifying their sustainability challenges, and informing on socioeconomic and ecological dimensions of groundwater. These insights have revealed groundwater as a dynamic component of the water cycle and social–ecological systems, leading to an expansion in groundwater science that increasingly focuses on groundwater's interactions with ecological, socioeconomic, and Earth systems. This shift presents many opportunities that are conditional on broader, more interdisciplinary system conceptualizations, models, and methods that require the integration of a greater diversity of data in contrast to conventional hydrogeological investigations. Here, we catalogue 144 global open access datasets and dataset collections relevant to groundwater science that span elements of the hydrosphere, biosphere, atmosphere, lithosphere, food systems, governance, management, and other socioeconomic system dimensions. The assembled catalogue offers a reference of available data for use in interdisciplinary assessments, and we summarize these data across their primary system, spatial resolution, temporal range, data type, generation method, level of groundwater representation, and institutional location of lead authorship. The catalogue includes 15 groundwater datasets, 23 datasets derived in relation to groundwater, and 106 datasets associated with groundwater. We find the majority of datasets are temporally static and that temporally dynamic data peak in availability during the 2000–2010 decade. Only a small fraction of temporally dynamic data is derived with any direct representation of groundwater, highlighting the need for greater incorporation of groundwater in Earth system models and data collection initiatives across socioeconomic, governance, and environmental science research communities. A small number of countries, led by the USA, Germany, the Netherlands, and Canada, generate most global groundwater data, reflecting a global North bias in the institutional leadership of these data generation activities. We raise three priority themes for future global groundwater data initiatives, which include: data improvements through prioritizing observed and temporally dynamic data; elevating regional and local scale data and perspectives to address challenges relating to equity and bias; and advancing data sharing initiatives founded on reciprocal benefits between global initiatives and data providers.

1. Introduction

Groundwater, a critical resource for drinking water, agriculture, and ecosystems, is under increasing pressure from human activities and climate change (Taylor *et al* 2013, Famiglietti 2014, Abbott *et al* 2019, Bierkens and Wada 2019, Gleeson *et al* 2020a, Scanlon *et al* 2023, Kuang *et al* 2024, Jasechko *et al* 2024, Reinecke *et al* 2024). Recognizing groundwater connections in social–ecological systems has been proposed as a holistic approach that makes possible a more complete assessment on groundwater's evolving role in the Anthropocene and supports more equitable groundwater management (Huggins *et al* 2023, Kuang *et al* 2024). Advancing this perspective requires a shift towards a broader system conceptualization where groundwater systems are understood as embedded within a network of socioeconomic, ecological, and Earth systems (Gleeson and Cardiff 2013, Huggins *et al* 2023) rather than in isolated contexts. This conceptual extension of groundwater systems to a more holistic social–ecological model is supported by both groundwater and social–ecological systems theory (Berkes and Folke 1998, Bloomfield *et al* 2008, Zellner 2008).

Indeed, there is a long history of scholarship on groundwater as a common-pool resource (Ostrom 1990, Blomquist 1994), and a rich literature details the benefits of applying social–ecological system framings to understand groundwater within economic, ecological, and governance contexts (Barreteau *et al* 2016, Rica 2017, Bouchet *et al* 2019). Whereas conventional groundwater investigations typically integrate hydrogeological, climatic, and topographic data (e.g. hydraulic conductivity, precipitation, and land surface elevation), assessments of groundwater in social–ecological systems require an integrated consideration of biophysical and social systems that extend beyond this scope to include data on ecosystems, governance, human activity, and the broader socioeconomic context (e.g. environmental flows, groundwater institutions, groundwater irrigation, and human development). Social–ecological framings center analysis on system interactions and thus show a strong potential to enhance understanding on groundwater's diverse roles across the Earth system (Gleeson *et al* 2020b), governance (Villholth and Conti 2018), human health (Wang *et al* 2023), food systems (Siebert *et al* 2010, Dalin *et al* 2019) and cultures (Re 2015, Zwarteveen *et al* 2021). Explicitly framed social–ecological system studies are beginning to populate the hydrology (e.g. Cabello *et al* 2015, Song *et al* 2024), groundwater (e.g. Lin *et al* 2020, Huggins *et al* 2024), and broader freshwater science (e.g. Carpenter *et al* 2015, Scown *et al* 2017, Varis *et al* 2019) literatures, yet our view is that these early studies only begin to realize the vast potential for further implementation of social–ecological freshwater science.

Systems connected to groundwater operate across a range of scales from the local to the global: stream-aquifer interactions at the reach scale (Brunner 2017, Yang *et al* 2025), basin scale management agencies and actions (e.g. Groundwater Sustainability Agencies in California, USA), groundwater-dependent ecosystems sustained by regional groundwater flow (Yao *et al* 2018, Aldous and Gannett 2021), transboundary aquifer governance (Shaminder and

Villholth 2017), international virtual water trade networks (Dalin *et al* 2017), and climate change drivers of regional groundwater storage (Wu *et al* 2020). In recognition of the myriad cross-scale interaction and broad interconnectedness between these processes, in addition to the widespread presence of groundwater challenges worldwide, a global research agenda on groundwater has emerged over recent decades (Konikow 2005, Giordano 2009, Foster *et al* 2013, Gleeson *et al* 2020a, Kuang *et al* 2024). This emerging focus on global-scale groundwater science has uncovered regional trends in groundwater storage (e.g. Richey *et al* 2015, Güntner *et al* 2024), the impacts of these trends on water security (e.g. Jasechko *et al* 2024), ecosystem health (e.g. Rohde *et al* 2024), and sea level (e.g. Wada *et al* 2012b), in addition to developing insights on large-scale interactions between groundwater and surface waters (e.g. Xie *et al* 2024), ecosystems (e.g. Sacco *et al* 2024), food systems (e.g. Dalin *et al* 2017, Scanlon *et al* 2023), and climate (e.g. Cuthbert *et al* 2019). However, the study of groundwater within holistic social–ecological system frameworks remains under-realized and few efforts to date explicitly integrate hydrological, social, ecological, and climate components into a global analysis (Bierkens 2015, Huggins *et al* 2024). Furthermore, as of yet there is an unquantified variety and volume of data to support such assessments.

Here, we conduct a review of global geospatial data available to study groundwater in social–ecological systems. Understanding this scope of data is a necessary but yet untaken step to support analyses on social–ecological and Earth system interactions with groundwater. These include systematic comparisons between regions to inform data generation and sharing priorities, characterize available data for evaluating and validating continental and global models, and facilitate global groundwater sustainability science (Famiglietti 2014, Gleeson *et al* 2020a, Scanlon *et al* 2023, Mukherjee 2024). Reviewing these global data may be further useful for non-global applications as global data can serve as a place-holder in regions where localized data are either unavailable or of insufficient quality.

This review seeks to complement several existing groundwater data sharing initiatives, such as the International Groundwater Resources Assessment Centre (IGRAC) Global Groundwater Information System (GGIS: https://ggis.un-igrac.org/), UNESCO's Intergovernmental Hydrological Programme's Water Information Network System (IHP-WINS: https://ihp-wins.unesco.org/), and the groundwater sector of the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP: www.isimip.org). While these initiatives share common goals of providing global groundwater and groundwater-related data to researchers, practitioners, and decision makers, none take an explicit social–ecological systems framing to the collection and provision of data. These initiatives tend to focus on groundwater resources (e.g. storage, recharge, aquifer boundaries) and infrastructure (e.g. managed aquifer recharge sites). They generally lack, however, associated, contextual data such as related ecological processes (e.g. groundwater-dependent ecosystems), social and economic settings (e.g. GDP, population, inequality), and collective action processes (e.g. governance, conflict). These existing groundwater data sharing initiatives remain important and offer valuable services to their user base, which our work complements and expands. Thus, we seek to fill a gap in these built-for-purpose data sharing initiatives by providing a cohesive account of a holistic array of social and ecological datasets useful for global groundwater science applications.

We focus our review on open access datasets to align our work with the open science movement and to actively encourage the continued shift towards open data in groundwater science. Open science is 'perhaps the most important paradigm shift in the recent history of scholarly publishing' (Clark *et al* 2021) through its democratization of information access and its promotion of transparency and reproducibility. Open data is particularly important to support equitable access to information that can be crucial to supporting improved decision making in resource-constrained regions. Thus, our review seeks to include and promote data that is both globally available and freely accessible to all.

Yet, globally available and freely accessible data does not eliminate the potential for bias. Clear biases towards overrepresentation of institutions or researchers from the global North have been identified in climate, environmental, and conservation sciences (Karlsson *et al* 2007, Hazlett *et al* 2020, Maas *et al* 2021), and for study areas of hydrological climate hazards (Stein *et al* 2024). Similar biases have been recently identified in groundwater modeling over model extents and with respect to non-local model development (Zamrsky *et al* 2025). However, these forms of bias have yet to be evaluated with respect to the generation of global groundwater data. This review therefore doubles as an opportunity to assess geographic trends in institutional data authorship. Understanding which regions are driving global groundwater data development, including the coordination of global data sharing initiatives, can be instructive to evaluate the representation of regional perspectives and needs in global initiatives. These insights may then be used as a basis to further explore opportunities and tensions between global groundwater data ambitions and local to regional datasets, priorities, and realities.

We refer to the compiled review of global groundwater datasets as a groundwater data catalogue (see results section 3.1). Upon compiling the catalogue, we address the research questions listed below. Building on the outcomes of these lines of inquiry, we conclude with a discussion on the needs and priorities for future groundwater data efforts based on identified limitations and opportunities.

- How many global datasets and dataset collections are openly accessible for studying groundwater in social–ecological systems?
- What is the distribution of datasets across different components of social and ecological systems?
- How many datasets within the catalogue can be considered temporally dynamic?
- What are the spatial (e.g. grid size or zonal unit) and temporal (e.g. time step) resolutions of each dataset, and how are these spatial and temporal properties distributed across the catalogue?
- How are datasets distributed based on the degree to which they explicitly represent, derive from, or are associated with groundwater in their data generation processes?
- What is the national distribution of institutional authorship of these datasets, and how does this distribution compare and relate to regional trends in groundwater challenges?

2. Review methodology

Our process to develop the global groundwater data catalogue involved: identifying and screening diverse data repositories and platforms to identify datasets for inclusion, establishing a protocol to guide evaluation of candidate dataset suitability for inclusion, specifying a theoretical framework to set the scope and structure the review, determining a set of dataset attributes on which to collect metadata , and generating internal review processes among co-authors. The general structure of this review is illustrated in figure 1.

We note that our focus on geospatial data presents unique challenges compared to standard literature review protocols, such as the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) (Page *et al* 2021). For instance, PRISMA and similar systematic review frameworks cannot easily be applied to geospatial data reviews as data are located across diverse sources, both within the literature and across various open data platforms, and because our focus herein is on synthesizing data availability and patterns in spatial dataset metadata rather than study findings. Thus, we developed a review protocol (figure 1(a)) that is flexible to accommodate these specific needs, that follows conventions established in complementary global geospatial dataset reviews that also focused on the integration of social and biophysical data (Lindersson *et al* 2020), and that incorporates elements of the PRISMA approach such as the sequential screening of data based on clear selection criteria. Below, we summarize the theoretical framing of our review scope (section 2.1, figures 1(b) and (c)), overview the databases and data platforms screened to identify candidate datasets for inclusion (section 2.2), describe the screening protocol to select datasets for inclusion in the catalogue (section 2.3), detail the metadata we collected on each dataset (section 2.4, figure 1(d)), and clarify dataset nomenclature specific to this study (section 2.5).

2.1. Framing groundwater in social–ecological systems

Social–ecological systems are defined as ‘complex, integrated systems in which humans are part of nature’ (Berkes and Folke 1998) and are composed of interacting social and ecological elements. ‘Social–ecological systems’ are a concept that has proven powerful in guiding research at the interface of human–environmental systems interactions and the governance and management of common-pool resources. Many frameworks exist to analyze social–ecological systems (Binder *et al* 2013), including the social–ecological systems framework (SESF; McGinnis and Ostrom 2014), but this study does not specifically adhere to any of them. Rather, we seek to support the broad application of social–ecological systems concepts, methods, and lines of inquiry to global groundwater.

To refine this basic conceptualization of social–ecological systems for the purpose of guiding our data review, we developed a composite approach using well-established Earth system and social–ecological system frameworks (figure 1(b)). To specify classes of biophysical systems, we used a common taxonomy of Earth systems that consists of lithosphere, hydrosphere, biosphere, and atmosphere components. For a review of groundwater interactions with Earth systems, we refer readers to Gleeson *et al* (2020b). To specify classes of social systems, we applied a groundwater-centric interpretation of the social elements of the SESF. This interpretation led to a specific focus on food systems, reflecting the significance of groundwater-agriculture interactions at the global scale. These include agriculture being the dominant driver of groundwater consumption globally, the importance of groundwater for irrigation supporting crop production and food security, the magnitude of groundwater embedded in international food trade (Siebert *et al* 2010, Wada *et al* 2012a, Dalin *et al* 2017), and the large volume of data on food systems in relation to other socioeconomic sectors and human dimensions of groundwater. Thus, the social system classes guiding our review consist of groundwater governance and management, food systems, and a broad and accommodating category of ‘other socioeconomic settings and sectors’.

Whereas previous studies have sought to solicit community opinion to identify priority variables to study and monitor social–ecological systems (e.g. Pacheco-Romero *et al* 2020), this is not the goal of our initiative. We refrain from such efforts for several reasons. The first is that we are strictly

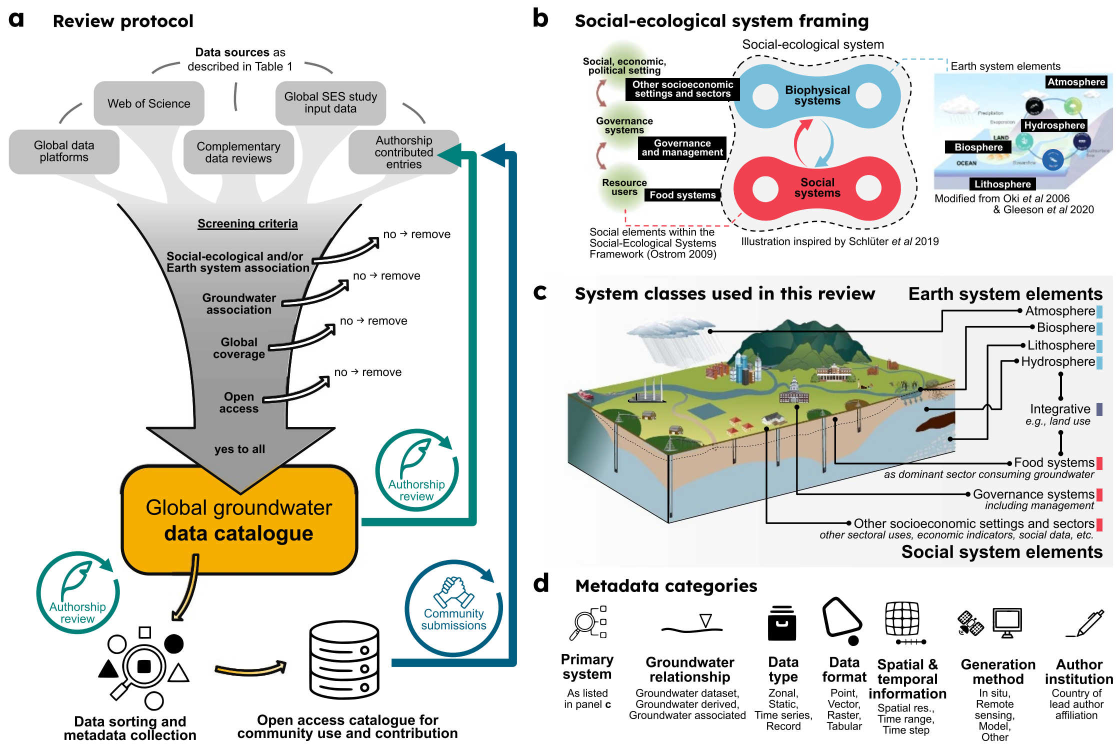

Figure 1. Review methodology. (a) Review protocol, which includes identifying potential source locations of candidate datasets, specifying explicit screening criteria to select and exclude datasets from inclusion in the data catalogue, and data sorting and metadata collection step, with study authorship review processes embedded within the protocol. (b) Conceptual framing of the review, which integrates Earth system and social–ecological system frameworks. (c) The system classes guiding the scope of data sought for inclusion in this review. (d) Metadata categories collected for each dataset and the possible values they can hold. Icons included in panel (a) are reproduced from The Noun Project (CC BY 3.0) and with permissions from Flaticon. The schematic in panel (b) draws inspiration from Ostrom (2009) and Schlüter et al (2019). This panel also uses an inset image modified from Gleeson et al (2020b), John Wiley & Sons. The central vector graphic in panel (c) is reproduced with permission from Huggins, X. © 2023 The Authors. Groundwater published by Wiley Periodicals LLC on behalf of National Ground Water Association (CC BY-NC-ND 4.0). Vector icons included in panel (d) are reproduced from The Noun Project (CC BY 3.0) and with permissions from Flaticon. See Open research section for vector icon attributions.

seeking to provide a data resource to facilitate a broader uptake of social–ecological research in global groundwater science rather than to shape the way in which groundwater within social–ecological systems is defined, characterized, and modeled. We anticipate that the research using this data resource will be richer, more diverse, and thus ultimately more useful than if constrained by a stricter system conceptualization. Secondly, developing a standard set of social–ecological system variables undermines the notion that social–ecological systems are place-based and that appropriate variables should be identified on a case-by-case basis.

Lastly, system boundary ambiguities are a characteristic aspect of engaging with social–ecological system framings (e.g. Andrachuk and Armitage 2015, Lazurko et al 2024), and we view the contestation of these boundaries to be a useful and insightful process. For instance, data on language diversity may offer implicit insights into the local complexity of groundwater governance and management or may provide a proxy representation of the diversity of value systems relating to groundwater. Others, however, may consider such data to hold little relevance for groundwater science. To this end, we hope this work's conceptual framing and the data included in our catalogue stimulate debate and generative reflections on the bounds that groundwater scientists set on their systems of study. In sum, our specified review scope (figure 1(b)) should be interpreted as an initial, community-based synthesis of the breadth of systems related to the study of groundwater in social–ecological systems.

2.2. Sources screened for dataset identification

We considered a diverse range of sources to identify candidate datasets for inclusion in this review. This broad search reflects the diverse locations where global geospatial data are hosted online. This approach is further necessary as the study of groundwater in social–ecological contexts lacks a broadly established self-identifying language or dedicated data repository.

Thus, we not only searched for datasets generated in publications, but also screened input datasets used in thematically aligned global social–ecological system assessments, leading global geospatial data platforms, compatible global data reviews, and through crowd-sourcing additional inputs from this study's co-authorship. The full list of consulted data sources and screening procedures are reported in table 1.

Table 1. Description of data sources consulted to develop our open data catalogue.

Data source category	Individual sources
Data used in global social–ecological system characterization studies	Ellis and Ramankutty (2008), Gain *et al* (2016), Sietz *et al* (2011), Václavík *et al* (2013), Varis *et al* (2019). →*All input datasets used in each study were screened.*
Global data platforms and compendiums	WRI Aqueduct (www.wri.org/aqueduct), WRI Resource Watch (https://resourcewatch.org/), IWRM data portal (https://iwrmdataportal.unepdhi.org/), IGRAC GGIS (https://ggis.un-igrac.org/), WWF Water Risk Filter (https://riskfilter.org/water/home), Protected Planet (www.protectedplanet.net/en), MapX (https://unepgrid.ch/en/mapx), GRID-Geneva data platform (https://unepgrid.ch/en/platforms), EarthStat (www.earthstat.org/), SEDAC (https://sedac.ciesin.columbia.edu/), Global Human Settlement Layer (GHSL) (https://human-settlement.emergency.copernicus.eu/datasets.php), Global Terrestrial Network—Hydrology (GTN-H) (www.gtn-h.info/), Copernicus Land Monitoring Service (https://land.copernicus.eu), Google Earth Engine Data Catalogue (https://developers.google.com/earth-engine/datasets/catalog), Open Land Map compendium (https://openlandmap.github.io/book/012-compendium.html). →*All datasets on each platform with global coverage were screened.*
Compatible global data reviews	Bolognesi *et al* (2018), Lindersson *et al* (2020), Kim *et al* (2021), Wang *et al* (2022). →*All datasets reviewed or summarized in each paper were screened.*
Web of Science search	Searches across the Web of Science core database were performed for the following query strings, and filtered using the 'associated data' tag: 'biophysical' *AND* 'global' *AND* 'dataset' (31), 'ecological' *AND* 'global' *AND* 'dataset' (255), 'governance' *AND* 'global' *AND* 'dataset' (36), 'groundwater' *AND* 'global' *AND* 'dataset' (44), 'socioeconomic' *AND* 'global' *AND* 'dataset' (95). Values in parentheses indicate the number of results for each query. All queries were performed in March 2024. →*All results from the above queries were screened.*

2.3. Dataset screening

We screened all candidate datasets for inclusion in our review using four explicit criteria: whether the data represent a variable with a clear social–ecological or Earth system association, whether the dataset is associated with groundwater, whether the dataset has global coverage, and whether the dataset is open access (figure 1(a)).

Our large authorship team of global groundwater, Earth system, and social–ecological system scientists hold a variety of perspectives that are reflected in these decision-making processes. All authors on this study, who represent a broad swath of expertise across multiple disciplines related to groundwater science, were able to contribute datasets, data platforms, and other sources to complement the review. We were not only interested in identifying data that have already been used to study global groundwater in social–ecological systems, but also in identifying data that have the potential for such use cases. Thus, we erred on the side of inclusion when making these screening decisions (see section 4.4 'review limitations' for further discussion).

Only open access datasets are included in the catalogue. We take this step to ensure the reviewed datasets are accessible for use in future studies, to encourage data sharing practices, and to broadly align this review with open science initiatives such as the Findable, Accessible, Interoperable, and Reproducible (FAIR) data principles (Wilkinson *et al* 2016) and the UNESCO Recommendation on Open Science (UNESCO 2021). In taking this step, this review implicitly evaluates the extent to which global groundwater data adhere to the FAIR principles of *findability* and *accessibility*. Screening for this criterion biases data generated within recent years (ca. 2015 and later), following the rise of open publishing and data deposition practices (Clark *et al* 2021, Hall *et al* 2022). The data sharing agreements for datasets that enable inclusion in this review include Creative Commons licenses, dataset-specific user agreements, or an explicit statement encouraging the use of data where a license or agreement was not readily identifiable. We reflect on the implications of this open data requirement in section 4.4.

2.4. Metadata categories

To review the catalogue, we derive a wide variety of metadata including: (1) the primary system to which the described dataset relates, (2) the way in which groundwater is related to the dataset, (3) dataset type and (4) format, (5) spatial resolution, (6) temporal range and time step for temporally dynamic data, (7) dataset generation method, and (8) institutional country of lead authorship (figure 1(d)).

The *primary system* to which the dataset relates reflects the individual systems included in our review scope, as presented in section 2.1. These seven system subclasses ('Atmosphere', 'Biosphere', 'Lithosphere', 'Hydrosphere', 'Food systems', 'Governance systems', and 'Other socioeconomic settings and sectors') are complemented by an eighth 'Integrative' class for datasets that describe variables which span multiple systems (figure 1(c)).

We categorized the *relationship* between groundwater and each dataset into three classes: groundwater data, groundwater derived data, and groundwater associated data. The 'groundwater data' class was assigned to actual groundwater data (e.g. water table depth, groundwater storage, or groundwater temperature), the 'groundwater derived data' class was assigned to datasets that explicitly incorporated groundwater consideration or representation in the data generation process (e.g. groundwater-driven wetlands and groundwater management indicators), and the 'groundwater associated data' class was assigned to the remaining datasets (e.g. cropland area, freshwater ecoregions, and gross domestic product). We applied a literal approach when assigning 'derived' versus 'associated' classes, where data were classified as 'associated' unless the data generation process included explicit representation of groundwater. Thus, this procedure to assign classes of dataset relationships to groundwater is not based on the strength of the underlying theory connecting a variable with groundwater but rather on the dataset generation process itself. For instance, global cropland datasets that do not indicate specific sources of irrigation water were classified as 'groundwater associated' despite agriculture being the dominant consumer of groundwater globally. However, should datasets explicitly consider groundwater, such as identifying areas equipped for groundwater irrigation, these data were classified as 'groundwater derived'. Similarly, global wetland maps that do not specify wetland type were identified as 'groundwater associated' whereas those that do specify groundwater-dependent wetlands were identified as 'groundwater derived'.

Dataset *types* represent the nature of the dataset as either zonal data (e.g. climate zones), a static dataset (e.g. farm field size for a given date), a time series (e.g. annual population estimates), or event or process records (e.g. freshwater treaties or acts of hostility). We assigned data types as 'zonal' if the principal use case is as a spatial unit for data summary (e.g. IPCC reference regions), and as 'static' if the primary use case is the documentation of an underlying system property or attribute (e.g. near-surface porosity) even if a secondary purpose of the data can be as a zonal layer. To differentiate between 'time series' and 'record' classes, we considered time series as data that are provided at regular time steps with a consistent spatial extent whereas historical record data typically have irregular time steps (e.g. water-related conflicts) with potentially inconsistent temporal ranges depending on individual entries within the dataset (e.g. water levels in monitoring wells). Should event records be synthesized into a dataset with regular time steps, these data would be recorded as a time series.

Dataset *formats* were assigned as raster, vector (e.g. polygon, polyline, or point data), or tabular. For raster data, we recorded the spatial resolution of the dataset. For vector data, we collected dataset-specific spatial information such as the median size of polygons or map scale based on metadata availability. For all temporally dynamic data, we recorded the start and end dates of the series along with the time step if data were available at regular intervals.

We additionally identified each dataset's *generation method* as being either (1) *in situ* observations,

(2) remote sensing observations, or (3) modeled or simulated data, such as datasets that have used statistical or process-based models to extrapolate data across larger domains, historical reconstructions or future projections, or approaches that combine observations with models to develop datasets for variables that are challenging to directly observe.

To assess the geographic distribution of *institutional authorship*, we recorded the country of the institution affiliated with each dataset's lead author. If corresponding and lead authorship differs for a dataset, we additionally included the location of the corresponding author's affiliation. For data with institutions as the data provider, we used the location of the institution's headquarters. When discussing results of these institutional distributions, we limited our analysis to only groundwater and groundwater derived datasets (i.e. excluding groundwater associated entries) to constrain insights and conclusions to the groundwater science community.

Similarly to the dataset identification and screening process, the assignment of metadata categories was reviewed by the study's co-authorship, reducing the potential bias of individualized decision making on study outcomes.

2.5. Dataset collection nomenclature

One challenging aspect of this review concerned how to best incorporate data from large, coordinated research communities such as: output from global hydrological models (Warszawski *et al* 2014, Schellekens *et al* 2017, Reinecke *et al* 2021), precipitation (Sun *et al* 2018), crop systems (Müller *et al* 2019), and other Earth observation datasets (McCabe *et al* 2017, Jaramillo *et al* 2024). These communities have respective data reviews and repositories (see preceding references), and including all associated datasets risked turning our exercise into an intractable 'review of reviews'.

To maintain our focus on reviewing the variety of data available to study global groundwater in social–ecological systems, we use the term 'dataset collection' to indicate when more than one dataset was included for a specific variable (e.g. precipitation). To ensure that dataset collections did not skew outcomes on data availability when summarizing across system types, we count dataset collections as a single dataset when reporting on the overall size of our catalogue (i.e. the size of our catalogue is reported as the sum of unique datasets and dataset collections).

We also use the term 'dataset collection' to represent data initiatives that collect a wide variety of variables within the same initiative. For example, the Worldwide Governance Indicators initiative develops six indicators of governance dimensions, yet all indicators are available over the same time range and the same spatial resolution, and are shared as a cohesive dataset. Rather than listing and reporting on these indicators individually, they are included in our review as a single dataset collection.

3. Results

3.1. A catalogue of open access datasets relevant to global groundwater

In total, our catalogue identifies and reviews 144 datasets and dataset collections available to study global groundwater in social–ecological systems (table 2). All datasets, including metadata and persistent weblinks, along with a searchable, interactive table with data access links, are provided in this study's data repository: https://github.com/XanderHuggins/groundwater-data-catalogue. We also intend for this catalogue to continue as a 'living resource', and a submission form is included on this online resource to allow for community contributions over time. Our intention is to periodically update the catalogue, subject to funding availability to continue this work.

Of the 144 unique datasets and dataset collections, 15 were classified as groundwater data, 23 as groundwater derived data, and the remaining 106 as groundwater associated data (figure 2(a)). Most groundwater datasets (e.g. depth to the water table, groundwater storage anomalies, groundwater temperature, etc) are classified in the hydrosphere category, whereas the 23 groundwater derived datasets are distributed across all categories but are most commonly represented in the biosphere (e.g. groundwater-dependent ecosystem extents), hydrosphere (e.g. water table ratio), and governance (e.g. management indicators) categories. We find most data with only groundwater associations within the 'other socioeconomic settings and sectors' category (24), which includes data such as population count, gross domestic product, and gender development inequalities. Patterns in groundwater relationships are evident and vary across the system classes. Half of all hydrosphere datasets are groundwater or groundwater derived datasets, whereas all other system categories are heavily skewed with groundwater associated data.

Data on 'hydrosphere' variables are the most common in the catalogue (figure 2(b)), with 36 datasets and dataset collections. Following the hydrosphere, the systems with the greatest variety of accessible datasets are 'other socioeconomic settings and sectors' (26), and the 'biosphere' (23). Conversely, 'lithosphere' (11), 'integrative' (9), and 'atmosphere' (8) have the least representation. 'Governance' (15) and 'food systems' (16) fall on either side of the median data availability per system class. These patterns in data accessibility may reflect the overall treatment and consideration of groundwater across research fields yet may also simply reflect the scope of individual system categories used in this review (e.g. the class 'other socioeconomic systems' is substantially broader than the more constrained 'atmosphere' or 'lithosphere' classes).

Table 2. Overview of unique variables and datasets included in the catalogue. The total counts of unique datasets and dataset collections are shown in figure 2(b).

System	List of variables * = part of a *dataset collection*
Atmosphere	**Records:** isotopes in precipitation, evapotranspiration observations* **Static:** aridity index **Time series:** precipitation*, extreme precipitation projections*, evapotranspiration*, hydrometeorological variable collections* **Zonal:** Köppen–Geiger climate zones, IPCC reference regions
Biosphere	**Records:** ecosystem fluxes, species abundances **Static:** environmental flow groundwater head limit, groundwater-dependent ecosystem extents*, amphibian and mammal species richness*, rarity-weighted species richness, ecological conservation prioritization index, global wetlands, wetlands of international importance, groundwater ecosystem biodiversity, ecohydrological classes of forest growth, ecosystem functional groups, root zone storage capacity and depth, groundwater-driven wetlands, soil organic carbon content, Ramsar wetlands **Time series:** vegetation indices* (e.g. NDVI, EVI), gross primary productivity*, maximum rooting depth, plant functional types, ecological vulnerability index, vegetation health index, wetland classification, plant functional types, dominant class of grasslands **Zonal:** freshwater and terrestrial ecoregions*
Lithosphere	**Static**: land subsidence, near-surface permeability and porosity, depth to bedrock, coastal aquifer thickness, soils*, lithological map, active faults, thickness of soil, regolith, and sedimentary deposits, crust model **Zonal:** karst aquifer map, sedimentary basin map
Hydrosphere	**Records:** river discharge, groundwater levels*, dam locations and metadata*, groundwater recharge, surface and groundwater salinity, global freshwater quality observations, karst spring hydrograph, isotopes in rivers **Static**: river width, streamflow indices*, groundwater response time, water table ratio, modern groundwater volume, river reach fragmentation, lakes*, lake bathymetry, lake volumes, river and stream intermittency, groundwater recharge*, terrestrial water storage rate of change, surface water extent*, groundwater vulnerability to floods and droughts, height above nearest drainage **Time series**: streamflow*, soil moisture*, water table depth*, terrestrial water storage anomaly*, drought indices*, groundwater temperature, groundwater storage anomaly* **Zonal:** watersheds, river network, aquifers, transboundary aquifers
Food systems	**Static**: crop allocation to end uses, gridded livestock systems*, crop harvested area*, crop type, crop production, crop yield*, field size*, cropland area*, planting and harvesting dates for major crops, pasture area, virtual water trade embedded in agriculture, area equipped for irrigation by source **Time series**: yield gaps, crop water footprints*, irrigation areas*, cropland extent*, crop yields*, harvested areas*, harvesting dates*, pesticide and fertilizer application rates
Governance, peace, management	**Historical records:** water related intrastate conflict and cooperation, international river basin organizations, international water events, international freshwater treaties, water conflicts, water related intrastate conflict or cooperation **Static**: environment, social, and governance (ESG) risk index **Time series**: Varieties of Democracy (V-Dem)*, Integrated Water Resources Management (IWRM) implementation indicators*, Worldwide Governance Indicators (WGI)*, environmental performance index, subnational corruption, World Values Survey* **Zonal:** administrative units, Indigenous territories, Indigenous treaties

(Continued.)

Table 2. (Continued.)

Other socioeconomic settings and sectors	**Records:** managed aquifer recharge schemes, living conditions of women and well-being **Static**: access to improved drinking water, roads*, power plants, accessibility to cities, development potential indices*, terrestrial human footprint, relative deprivation index, travel time to healthcare **Time series**: freshwater withdrawal by sector, human modification of terrestrial lands, human footprint*, gross domestic product (GDP)* and GDP per capita, population*, migration*, urban land*, human development index, electricity consumption, nighttime lights*, migration*, gender development inequality, social adaptive capacity, gini index, GNI per capita, human development, gender inequality, inequality databases* **Zonal:** Indigenous languages, protected areas and other effective area-based conservation measures (OECMs)
Integrative	**Records:** World Bank DataBank indicators*, AQUASTAT core database* **Static**: population distance to surface freshwater, land use decision making archetypes, land system archetypes, forest and tree proximate people, human appropriation of net primary productivity **Time series**: land cover*, land use change*, river basin resilience, anthropogenic biomes*

Note: Bold text signifies the 'data type' groupings of datasets included for each system.

We find several patterns within the dataset format, type, and generation method categories across the system classes (figure 2(b)) and which generally reflect methodological differences across disciplines that relate to groundwater. For instance, the majority of governance data is tabular, consists largely of event records (e.g. treaties, conflicts, etc), and is derived through means outside of *in situ* observations, remote sensing, and models. Conversely, biophysical data are predominantly raster data, split between static and time series formats, and largely derived through models. Socioeconomic data are predominantly temporally dynamic, whereas the lithosphere datasets are entirely zonal or static.

We find these datasets are most commonly provided at moderate spatial resolutions between 5 and 30 km (figure 2(c)). Biosphere data are most common among datasets available at very fine resolutions (<1 km), while socioeconomic data are most commonly available at fine resolutions (1–5 km), and food systems at moderate resolutions. Fine resolution data area most common among hydrosphere datasets, while governance data are most often available at the national scale.

3.2. Time series availability and groundwater underrepresentation

We compared temporal ranges of all temporally dynamic datasets, regardless of time step, across system type, spatial resolution, and relationship to groundwater (figure 3). In this analysis, we included record datasets alongside time series datasets when temporal ranges are reported, even if all record locations within the dataset are not available over the full reported range. We find that the greatest overlap among temporally dynamic data occurs over the 2000–2010 decade, peaking with over 80 available datasets. A considerable spike in time series data occurred in 1990 and a consistent decline in time series availability is visible since ∼2015. This creates an uncomfortable reality and priority for on-going and future data initiatives: until subsequent datasets are published or updated, there is a greater volume of temporally dynamic data over the 2000–2010 decade than 2010–2020, even though we are now well into the third decade of this century.

'Other socioeconomic settings and sectors' represent the most common class with temporally dynamic data, whereas zero dynamic lithosphere datasets were identified (figure 3(a)). Moderate spatial resolutions are the most common within temporally dynamic datasets (figure 3(b)). Further, we can observe the emergence of datasets with very fine spatial resolutions in the 1990s. We find very few temporally dynamic groundwater and groundwater derived datasets and thus most time series in the catalogue are groundwater associated datasets (figure 3(c)). Only a small subset of time series datasets or dataset collections can be considered actively updated through ongoing efforts.

3.3. Institutional leadership patterns reveal a global North bias

A small set of countries lead the development of these global datasets. Among groundwater and groundwater derived datasets ($n = 38$), 10 were led by institutions located in the USA, 8 by institutions located in Germany, and 5 by institutions located in Canada and the Netherlands, respectively (figure 2(d)). Only 13 countries have led the development of all groundwater and groundwater derived datasets (and only 19 countries if extending to all datasets included in

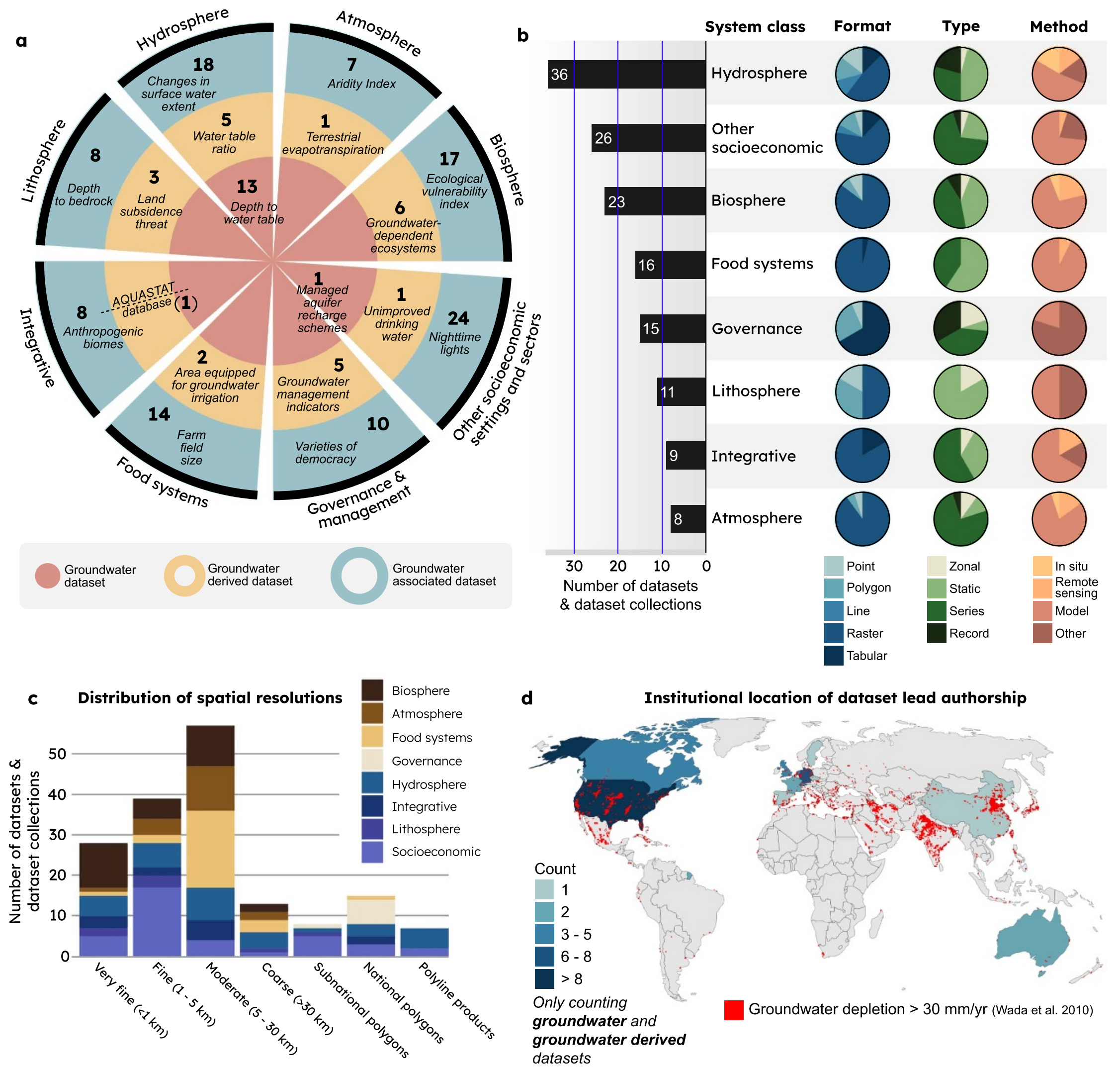

Figure 2. Distribution of reviewed datasets across metadata categories. (a) Dataset count per system class and category of groundwater relationship. Example datasets of each class are provided (e.g. *Nighttime lights* for a groundwater associated dataset in the 'other socioeconomic setting and sectors' class). (b) Histogram: counts of unique datasets and dataset collections per system class. Pie charts: distributions of metadata categories per system class for all datasets in the catalogue. (c) Distribution of spatial resolutions across system classes for all datasets in the catalogue. (d) Distribution of institutional authorship locations compared with regional hotspots of groundwater depletion (Wada *et al* 2010).

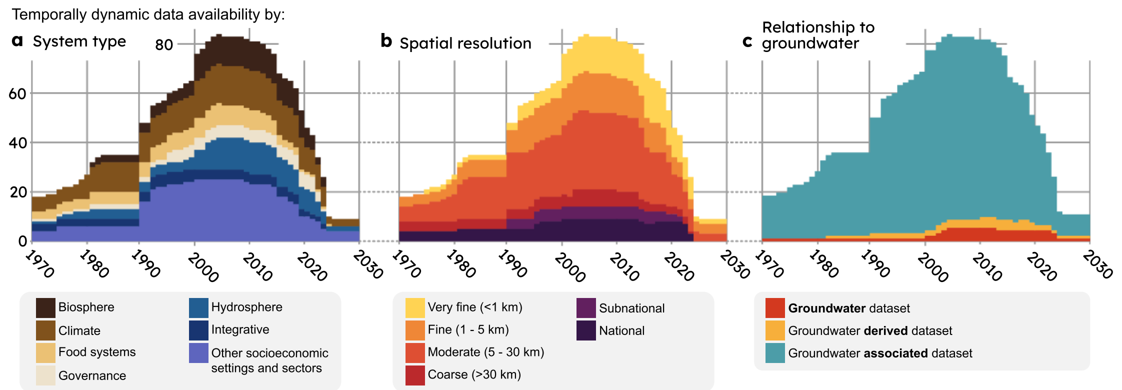

Figure 3. Availability of temporally dynamic datasets. (a) Dataset availability by system type. (b) Dataset availability per spatial resolution. (c) Dataset availability per groundwater relationship category. These plots are generated using the first and last years of available data per dataset and do not represent the time steps of individual datasets. These plots do not group dataset collections together and thus represent temporally dynamic data availability across all datasets in the catalogue.

our catalogue). International agencies are responsible for 8 of the groundwater and groundwater derived datasets and are thus not shown in the inset map in figure 2(d). We note that these totals do not sum to the number of groundwater datasets as two countries are counted if lead and corresponding authorship institutional affiliations vary. Several countries experiencing severe groundwater sustainability challenges stemming from groundwater depletion, including India, Pakistan, Iran, Mexico, and Japan (figure 2(d)), are absent from the institutions generating these global groundwater datasets. Countries in South America and Africa are also absent from the lead authorship of these global data products, and Australia is the only country represented from the southern hemisphere. Together, this suggests that the processes of generating these global datasets have been overwhelmingly led by institutions from the global North. We reflect on the implications of this reality in section 4.3.

4. Discussion

Drawing from the above review of this global open data landscape, we identify three priority themes for global groundwater data: embracing a social–ecological research agenda for groundwater (section 4.1), addressing gaps and limitations in global data products (section 4.2), and implementing more equitable and representative processes (section 4.3) (figure 4). These themes correspond with core opportunities and challenges identified in the global data landscape, provide actionable recommendations for the research community, and are oriented in the spirit of ensuring that global groundwater science and data remain use-inspired and responsive to the evolving needs of researchers, decision-makers and practitioners.

4.1. Existing richness of data with significant potential for synthesis and analysis

Our assembled global groundwater data catalogue depicts a portrait of the open data landscape supporting the study of groundwater in social–ecological systems that is large and diverse. Although not formally evaluated by our methodology, our authorship team of global groundwater, Earth system, and social–ecological system scientists shares the perspective that only a small portion of the data included in our assembled catalogue has already been implemented in groundwater-focused studies. In our view, this points to the significant and unrealized potential for social–ecological system mental and conceptual models, methods, and research objectives in large-scale groundwater science. Before we discuss clear opportunities for improvement in this data landscape (section 4.2), we first seek to emphasize that the research community does not need to wait for greater or improved data availability: rich data are already available. Indeed, combining already existing data may prove to be effective and capable of uncovering new and important relationships between groundwater and social–ecological system elements and processes, and in guiding the identification of which social–ecological data we most urgently lack.

We recognize that embracing interdisciplinary, social–ecological forms of analysis may be challenging to groundwater scientists as this requires the integration of data, concepts, and methods outside of the scope of conventional hydrological science and training. This shift, however, is not the sole burden of groundwater scientists, as social–ecological and Earth system scientists equally are called to place a greater emphasis on groundwater systems and groundwater dynamics. On this basis, the development of a community-wide research agenda, similar to what has been done in the hydrological community (see Blösch *et al* 2019), and a methodological handbook for groundwater analysis in social–ecological systems, following the example of broader social–ecological systems scholarship (Biggs *et al* 2021), could be two highly valuable contributions to guide this shift. While our hope for this review and the global groundwater data catalogue is to further establish the conditions for this shift to occur, outlining this prospective research agenda is outside the scope of this work.

We view the concept of essential variables, such as pursued in the fields of climate (Bojinski *et al* 2014), biodiversity (Pereira *et al* 2013), oceans (Miloslavich *et al* 2018), and ecosystem services (Balvanera *et al* 2022), as one compelling way to integrate a coherent social–ecological framing of groundwater systems with global data and analysis. Essential variables aim to identify a necessary set of variables to sufficiently monitor and detect changes in the function and structure of a given system. In the essential climate variables (ECVs) initiative, groundwater is directly included as an ECV and is implicitly represented in the terrestrial water storage variable (GCOS 2024). Yet, we foresee the potential for a broader and dedicated set of essential variables for groundwater in social–ecological systems, and which could include a subset of the datasets included in the global groundwater data catalogue.

We do not seek to establish a list of essential groundwater variables here as this would necessarily involve a community-wide, iterative, and engaged process. However, we offer a starting point by considering the potential of the groundwater essential variable concept. Such an initiative could identify a fundamental and coherent set of groundwater system properties and functions in social–ecological systems that have, or could obtain, observational monitoring or reporting capacities. Further, the initiative could serve the need to harmonize these data into analysis-ready formats, and more broadly could act as a vehicle to organize, develop, synthesize and

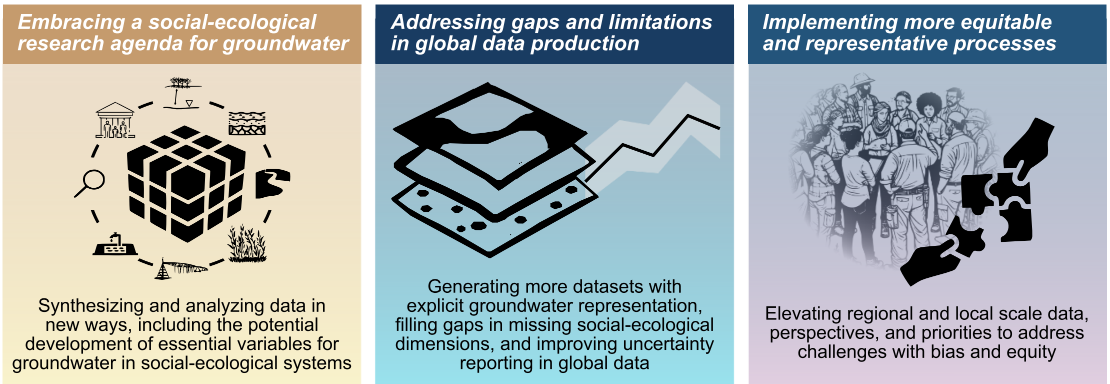

Figure 4. Identified priority themes for global groundwater data and science in social–ecological systems. Icons within this figure are reproduced from The Noun Project (CC BY 3.0) and from the Integration and Application Network (CC BY-SA 4.0). See open research section for vector icon attributions.

fund global initiatives on the study of groundwater in social–ecological systems.

4.2. Addressing data limitations including groundwater representation, uncertainty, and blind spots in global data

Our findings also suggest that substantial needs exist to improve this data landscape because all system classes outside of the hydrosphere skew towards datasets only associated with groundwater, and as datasets derived in relation to groundwater only account for a small fraction of temporally dynamic datasets. Thus, while there exists substantial potential to apply already available data, the question arises: how can the global-scale groundwater data community make concerted efforts to address these limitations and generate a more extensive and capable global groundwater data landscape?

Using the various classification schemes implemented in our review, we can identify a set of preferences for future data efforts that include: observed over modeled data, time series over static datasets, and groundwater derived over groundwater associated data. For instance, datasets on the extents of groundwater-dependent ecosystems and areas equipped for groundwater irrigation are currently available for specific time slices but could offer a myriad of potential insights if both datasets were generated over consistent time ranges (i.e. improving static datasets to time series). There is initial progress in this direction with the generation of a dataset on the temporal evolution of irrigated areas from 1900 to 1980 in 10 year time steps, and from 1980 to 2015 in 5 year time steps (Siebert *et al* 2015, Mehta *et al* 2024). However, similar improvements are yet unrealized for groundwater-irrigated areas.

Improving temporal frequency and aligning time steps across existing datasets would significantly improve the scientific potential of these data and would ease the integration of multidimensional data into analysis frameworks. These improvements stand alongside our finding that temporally dynamic data have declined in availability since 2000–2010, highlighting the need for substantial efforts to restore past levels of data availability and ensure sustained data coverage in the decades ahead. We do not necessarily attribute this to a decline in global time series data generation as a time lag is necessary for research efforts to synthesize and publish data covering recent years, particularly for variables and processes that do not benefit from near-real time observational capacities. Reduced incentives, perceived or real, to update and extend existing datasets relative to the incentives of publishing a dataset that is a 'first of its kind' may also contribute to the observed decline in data availability. Thus, it may be beneficial for groundwater-related societies, journals, and funding agencies to reflect on potential initiatives that can create incentive structures to equitably reward dataset updates alongside original dataset development.

Without sufficient temporally dynamic data, testing hypotheses on dynamic social–ecological system behaviors of groundwater systems such as emergence, regime shifts, context dependence, and system resilience (e.g. Preiser *et al* 2018) may be limited to conceptual and theoretical realms (Di Baldassere *et al* 2015, Troy *et al* 2015). These limitations not only create barriers to scientific inquiry but can more problematically impede understanding of complex system dynamics and contribute to erroneous decision making in applied contexts (Chávez García Silva *et al* 2024).

Enhancing the availability of groundwater derived datasets across social–ecological system features is not trivial and will require a variety of approaches. These approaches could include: improved groundwater representation in global Earth system, crop, and dynamic vegetation models; a greater focus on groundwater systems in global applications of statistical and machine learning models; the development of emerging modeling approaches, such as agent-based models, that incorporate a groundwater component; and the collation of regional datasets into global repositories (see section 4.3).

The present lack of globally distributed temporally dynamic datasets may point to a future of global groundwater science that is more oriented towards case study and point location-based analyses. These initiatives (e.g. Kreibich *et al* 2023, Tiwari *et al* 2023) may more readily be able to implement existing observational capacities, and may more vividly reflect contextually rich data, such as dimensions of human health or ecosystem services that are challenging to organize into globally distributed datasets that often require common conceptual models and methodologies for monitoring, extrapolation, or modeling over the global domain. Thus, an intermediary level of analysis consisting of globally distributed case studies integrating groundwater with social–ecological data may offer pragmatic and instructive insights on both the dynamics of groundwater in social–ecological systems, and to guide future global data initiatives.

There is also a need for data development on currently missing or underrepresented social–ecological system elements associated with groundwater. For instance, datasets connecting groundwater with domestic use and human health (e.g. Mukherjee *et al* 2019) such as health outcomes linked to groundwater salinization (Mueller *et al* 2024), observations on interconnections between groundwater and surface water (e.g. Jasechko *et al* 2021), and that capture human factors such as behavioral (e.g. Castilla-Rho *et al* 2017), economic (e.g. Bierkens *et al* 2024), infrastructural, legal (e.g. Nelson and Perrone 2016, Rohde *et al* 2017), institutional, and governance (e.g. Villholth and Conti 2018) relations to groundwater are particularly rare in the literature. These data gaps likely exist due to a combination of methodological and disciplinary challenges. For instance, processes to standardize and collect data representing context-sensitive dimensions of water governance and other forms of collective human behavior are more contested and challenging compared to biophysical variables. Disciplinary fragmentation may also contribute to these data gaps as related research fields may not be immediately aware of the benefits of incorporating groundwater interactions into their research programs and related datasets. The onus to change this awareness lays in part on groundwater scientists to more vocally articulate the reality that physical sciences provide only a partial understanding of groundwater dynamics and thus offer necessary but insufficient knowledge for effective and equitable governance and management. Prioritizing inter- and transdisciplinary research team composition is one pragmatic strategy to overcome the fragmentation that contributes to persistent data gaps, and which fosters the collaborations needed to address them.

Several other foundational groundwater datasets would benefit from continued improvement from their original releases. For instance, the development of comprehensive and harmonized geological datasets that combine global lithology maps with global borehole records would enable a cascade of wider dataset improvements including more reliable global groundwater models, improved representation of groundwater in Earth system models, and a strengthened ability to convert observations in groundwater storage to changes in the water table.

Finally, we raise the need for more robust inclusion and reporting of uncertainty in global datasets (Wagener *et al* 2021). Without uniformly enforced and followed practices of uncertainty reporting in the development of global datasets, data selection for applied studies can often be driven by operational convenience, such as ease of integration based on spatial or temporal resolutions, rather than a critical evaluation on the implications of the sourced data's uncertainty. Given social–ecological assessments inherently combine a wide variety of data, reporting on uncertainty becomes all the more important as compounding uncertainties can have important implications on study outcomes that may lead to erroneous or tenuous decision making. Integrating uncertainty in social–ecological data presents additional challenges as uncertainty is often reported in different ways across natural and social sciences (see Westerberg *et al* 2017). These challenges are not unique to the study of groundwater in social–ecological systems but will be important to address to ensure this interdisciplinary research direction for groundwater science is rigorous, reproducible, and relevant for applied use cases.

4.3. Elevating and respecting regional and local perspectives, priorities, and needs in global initiatives

An abundance of regional and local scale data is missed in this review that is focused on global data. These range from large, nation-scale initiatives on monitoring wells, groundwater well uses (Lin *et al* 2024), sub-national virtual water flows (Dang *et al* 2015), to a myriad of crowd-sourced data within individual aquifers and basins. Indeed, the volume of global data accumulated in our reviewed catalogue is but a small fraction of the total volume of data that may exist to understand and manage groundwater. Groundwater data needs vary substantially, mirroring the diversity of geographies, ecologies, and societies within which groundwater is situated and connected. Local and regional datasets will inherently correspond better to these needs than those generated through global initiatives. As we focus below on potential avenues to integrate these data into global initiatives, we first emphasize that local and regional data are essential, relevant, and valuable regardless of their inclusion in global initiatives.

Groundwater and groundwater-connected systems operate across a range of scales, and integrating data generated across a variety of scales is one opportunity to better reflect multi-scalar processes in global datasets. Existing data 'networks of networks' such as the Global Groundwater Monitoring Network (IGRAC 2025) and FLUXNET (Delwiche *et al* 2024), and community initiatives such as the Groundwater Model Portal (Zipper *et al* 2023) demonstrate the potential for bottom–up collaboration on global dataset development yet substantial work lies ahead to realize this potential across a wider set of social–ecological dimensions. Involvement of non-governmental organizations, global development organizations, and intergovernmental organizations could play an instrumental role in providing incentives, investment, infrastructure, and/or enforcement (e.g. through strong data mandates) of engagement in global initiatives. However, development of such initiatives is an inherently political process whose success will depend on equitable stakeholder and rightsholder engagement, transparent data-sharing frameworks, and the challenge of ensuring mobilized funding is allocated based on needs (see Stein *et al* 2024).

While synthesizing existing regional datasets and initiatives is a laudable goal, it is also one accompanied by substantial challenges. Our observation that global groundwater data are generated and studied in institutions mostly distant from the regions facing acute groundwater sustainability challenges raises important questions about what research priorities are driving global scale research and whose interests these priorities primarily serve.

Our aim in highlighting the geographical centers of global data leadership is to encourage reflection on the potential implications of their concentration in the global North. For instance, recognition of the underrepresentation of scientists from tropical Africa in global climate research has, over the last decade, led to more inclusive research and improved human and infrastructural capacities (e.g. Senior *et al* 2021, Lamptey *et al* 2024) that are now addressing long-held gaps in tropical meteorology and the representation of Africa's monsoon-dominated climatology in climate projections (Senior *et al* 2021, WMO 2022). Similarly, which groundwater connections and uses might we, as a global groundwater community, be underrepresenting in our initiatives?

We direct this discussion toward scientists from overrepresented institutions in high-income and historically dominant regions that are largely removed from the acute and localized challenges of groundwater sustainability (figure 2(d)). Our aim is not to place additional expectations on scientists in underrepresented or under-resourced regions to engage more frequently in global initiatives, particularly in contexts where local and regional challenges demand urgent attention and represent more immediate priorities. Rather, we emphasize the need for more inclusive and co-led processes that meaningfully involve academics, practitioners, and partners from across under-resourced contexts. Yet, achieving this without reinforcing existing asymmetries in power and authority remains a complex but necessary challenge to confront.

Data sharing and equitable participation of underrepresented institutions from developing regions are essential components of global data initiatives. However, engagement between local, regional, and global initiatives needs to address long-standing challenges, such as research imbalances and the ownership of scientific advances that are undermined by 'helicopter' research (the practice of researchers from well-resourced institutions and regions conducting work in under-resourced regions without meaningful involvement or leadership from local collaborators) (Gbondo and Michelsen 2024). Developing global data sharing initiatives on the principle of reciprocal benefits will be essential but alone may not be sufficient to address these challenges. Potential benefits to reciprocate the sharing of data from researchers and national or regional agencies may include increased visibility of data and research (such as through co-authorship on derivative studies), enhanced access to technological capacities for data processing (such as conversion of unprocessed data to standardized formats, data cleanup, and statistical summarizing), the inclusion and citation of data in global initiatives (e.g. WMO 2024), and the elevating of regional groundwater challenges and solutions to a broader, global audience. However, such benefits will need to be established on a case-by-case basis, tailored to the specific needs of individual data providers, including maintaining the ability of place-based communities, such as Indigenous Peoples, to control data access and the authority to ensure the stories told through these data are consistent with community-specific needs and values. These initiatives may require financial commitments for data access to sustain groundwater monitoring programs previously reliant on data access paywalls.

While the above discussion addresses representational biases in global groundwater data generation, it does not resolve broader ethical tensions between Big Data, open science, and Indigenous Data Sovereignty principles (Carroll *et al* 2020). Given that as much as 65% of land area is held under Indigenous Peoples' and local community customary systems (RRI 2015), these are truly global ethical questions and priorities for land-based sciences (Meyfroidt *et al* 2022). In settler-Indigenous contexts and beyond, it is important to ensure that creating open global datasets does not violate the privacy and security of individuals and communities by sharing sensitive information

(Zipper *et al* 2019) nor violate Indigenous data sovereignty (Walter *et al* 2021).

Even seemingly innocuous global data such as the type and presence of groundwater-dependent ecosystems may contradict Indigenous Peoples' authority to control data access that follow best practices established in recent data ethics frameworks (Carroll *et al* 2021). Indeed, the Indigenous territories data layer included in our review may be useful for the global groundwater community to better consider how Indigenous issues, priorities, land, and knowledge systems interact with one's work and may inform how engagement with Indigenous peoples may be necessary to meet ethical standards. It therefore is critical to clarify how the FAIR and CARE (collective benefit, ability to control, responsibility, and Indigenous ethics) data principles (Wilkinson *et al* 2016, Carroll *et al* 2020) interrelate and relate to global groundwater research.

4.4. Review limitations

Our interest in performing a wide sweep of social–ecological systems data for groundwater science applications is a process that requires normative judgments on which data and variables are considered as associated with groundwater. In acknowledgement of this, we characterize the presented groundwater data catalogue as a large, broadly representative, but not exhaustive resource. We anticipate that the primary benefit of this initiative will be to serve as a resource for scholars, practitioners, and others to identify data for use in future studies of groundwater in social–ecological systems. We note again our intention for the data catalogue to continue as a 'living resource', and future community contributions will further enhance the comprehensiveness and value of the catalogue.

The conceptual model that guided our study (figure 1(c)) centers social–ecological system dynamics in relation to shallow (i.e. upper ∼100 m of the subsurface) and terrestrial groundwater systems, yet underrepresents offshore aquifers, deep groundwater systems, and geochemical data. These biases can be addressed through complementary initiatives or may be addressed through future community contributions. While some geochemical datasets are included in our data catalogue, they are principally oriented around human health implications. Thus, our catalogue does not serve as a review of groundwater quality and geochemistry datasets (Misstear *et al* 2023), nor does it cover regional groundwater datasets that serve important roles in validating global models. Both of these topics warrant dedicated reviews.

We also note that while our commitment to open access data aligns our initiative with the open science movement, it also introduces additional biases. The open availability of data does not necessarily imply these data can uniformly be used as analysis-ready, 'off the shelf', products. Further development of initiatives such as FLUXNET (Delwiche *et al* 2024), that have established standardized protocols, formats, and quality control procedures, would support a broader shift towards analysis-ready data across diverse components of social–ecological systems. However, as already noted, developing such standardized protocols for qualitative, normative, and contested forms of non-biophysical data may prove more challenging.

Indeed, evaluations on data quality, dependencies, and provenance are all necessary before using open datasets in analysis. For instance, shared dataset dependency can prevent or limit data integration in system interaction studies (Lindersson *et al* 2020). Many datasets also contain specific user guidelines, such as water storage datasets specifying minimum resolutions of analysis despite source data being provided at finer scales (www2.csr.utexas.edu/grace/RL06_mascons.html). Our open data catalogue does not alleviate these requirements for data users.

Secondly, and as noted in section 4.3, open data initiatives face geopolitical barriers to data sharing and technical capacity gradients that can produce representation biases. We thus reemphasize that open data availability is a subset of overall data availability and that gaps in open data should not be assumed to simply reflect a lack of data. For instance, while platforms such as the GGMN may suggest limited data across most of Africa, data do exist in many regions but are inaccessible or not shared. These geopolitical barriers and technical capacity limitations remain on-going, and in some cases increasing, challenges to the open science movement and offer a reminder that data generation and sharing are both scientific and political processes. The practical implication of our commitment to open data is that proprietary, paywalled, or otherwise inaccessible data are not summarized in our review and are not included in the data catalogue. While this introduces a selection bias to our study, we believe this approach is merited given the clear benefit that all datasets in the catalogue are accessible to anyone.

Lastly, our analysis of institutional authorship only focused on the affiliations of lead and corresponding authors and did not analyze the full diversity of co-authorship lists. Many of these datasets were developed in large author studies and often do include institutions from lower-resource countries underrepresented in lead authorships. While we view this approach as a pragmatic account of the dominant geographies and institutional contexts of these data generation activities, future work would be welcomed that additionally analyzes the distribution of full authorship lists. Additionally, it would be valuable to analyze the extent to which local-scale data are integrated into global data products and to review the mechanisms through which data providers are acknowledged and included in global data synthesis initiatives.

5. Conclusion

To support continental- to global-scale research on groundwater in social–ecological systems, we developed and reviewed a large catalogue of global datasets and dataset collections ($n = 144$) that relate to groundwater systems and their social–ecological system interactions. We reveal that a rich variety of data are available for implementation in global studies, and our catalogue can serve as a reference for researchers to locate sources of interdisciplinary and open access data. We find important limitations and biases in the existing data. At the forefront of these limitations is a lack of temporally dynamic datasets that are explicitly derived in relation to groundwater, undermining the ability of global groundwater science to generate a strong evidence base for social–ecological system dynamics. A clear bias in institutional leadership from the global North is identified in these groundwater data generation efforts which prompts questions about the potential mismatches in needs, interests, and incentives between groundwater data generation and groundwater data needs.

Three priority themes for the global groundwater research community are identified: leading and innovating on new social–ecological forms of groundwater analysis, improving the data availability landscape through targeted dataset generation to fill gaps and limitations, and an increased focus on equitable and representative data and research processes. These core themes distill a number of actionable recommendations, including:

- Develop a community-wide **research agenda** and methodological approaches for studying groundwater in social–ecological systems.
- Clarify how **ethical frameworks** of open data (FAIR) and Indigenous data (CARE) principles interrelate and guide global **groundwater data governance.**
- Establish **reciprocally beneficial** initiatives to enable **regional data collation and synthesis** in cohesive global frameworks.
- Consider establishing **essential groundwater variables**, to guide social–ecological system conceptualizations and enhance the monitoring of groundwater's social, economic, and ecological functions.
- Improve the representation and integration of **groundwater in modeling frameworks** (e.g., hydrological, Earth system, agent-based, system dynamics, integrated assessment, and machine learning models).
- **Restructure incentives** to encourage **targeted data generation** on missing variables, promote the updating of datasets following their original publication, and support the creation of quality controlled, analysis-ready data products.
- **Mandate** the reporting of **data uncertainty** in peer reviewed data repositories.

These recommendations call on a variety of actors: researchers, to engage more deeply in interdisciplinary groundwater assessments and to collaborate more inclusively with global co-authors; journals and data repositories, to require uncertainty estimates in geospatial data and to give equal value to dataset updates as to original data submissions; funding bodies, to support large-scale groundwater research and data development; and international agencies, to invest in, incentivize, and build infrastructure for equitable global data sharing. Taking these steps will help ensure that groundwater science remains a use-inspired discipline that is responsive and relevant to the evolving demands presented by global environmental and social change.

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: https://github.com/XanderHuggins/groundwater-data-catalogue. The repository's readme file includes an access link to a searchable, online version of the data catalogue.

Acknowledgment

XH acknowledges a fellowship from the Natural Sciences and Engineering Research Council of Canada (NSERC) and a financial contribution from the Canadian Space Agency. TW acknowledges support from the Alexander von Humboldt Foundation in the framework of the Alexander von Humboldt Professorship endowed by the German Federal Ministry of Education and Research (BMBF). RGT acknowledges a fellowship (Number FL-001275) from the Canadian Institute for Advanced Research (CIFAR) under the Earth 4D: Subsurface Science and Exploration programme. MK acknowledges support from the European Research Council (Grant No. 819202), Research Council of Finland's Flagship Programme through project Digital Waters (Grant No. 359248), and Strategic Research Council through project Water&Food (Grant No. 365512). MP acknowledges support from the European Research Council (Grant No. 101118083) and the Strategic Research Council at the Research Council of Finland (Grant Nos. 365752, 365754). GDB acknowledges support from the TRANSCEND project funded by the European Commission (Horizon Grant No. 101084110). SL acknowledges support from the Centre of Natural Hazards and Disaster Science (CNDS), Sweden. IDG acknowledges support from the European Research Council (Grant No. 10104110). ML acknowledges support from the Taiwan NSTC (Grant 112-2111-M-002-021). YP acknowledges support from the U.S. National Science Foundation (BLUEGEM;

Grant No. 2127643). MFPB acknowledges support from the ERC Advanced Grant scheme (Grant No. 101019185—GEOWAT).

Open research

All analyses were conducted using the R project for statistical computing (R Core Team 2023), using the packages *tmap* (Tennekes *et al* 2018), *ggplot2* (Wickham 2016), and *MetBrewer* (Mills 2022). Composite figures were assembled in Affinity Designer (https://affinity.serif.com/en-us/designer/).

Figures 1 and 4 use the following icons from Noun Project (https://thenounproject.com/): 'author' by Ahsan Damar, 'classification' by Sinta Maulana, 'collaboration' and 'support' by Andika Cahya Fitriani, 'grid' by Yosua Bungaran, 'satellite' and 'authorship' by Lufti Gani Al Achmad, 'soil' by Candy Design, 'river' by Friedrich Santana, 'plants' by Hawraa Alsalman, and 'groundwater' by Adrien Coquet; 'Server' and 'Archive' free icons from Flaticon.com; and the 'Center pivot irrigation' vector illustration by Kim Kraeer and Lucy Van Essen-Fishman from the Integration and Application Network (https://ian.umces.edu). Other vector images incorporated in figures were either generated through generative AI (chatGPT by OpenAI) with prompts from the author or were drawn manually in Affinity Designer.

ORCID iDs

Xander Huggins 0000-0002-6313-8299
Tom Gleeson 0000-0001-9493-7707
James S Famiglietti 0000-0002-6053-5379
Robert Reinecke 0000-0001-5699-8584
Daniel Zamrsky 0000-0001-6046-688X
Thorsten Wagener 0000-0003-3881-5849
Richard G Taylor 0000-0002-9867-8033
Megan Konar 0000-0003-0540-8438
Claudia Ruz Vargas 0000-0002-4698-3685
Miina Porkka 0000-0002-8285-6122
Lan Wang-Erlandsson 0000-0002-7739-5069
Inge de Graaf 0000-0001-7748-868X
Mark Cuthbert 0000-0001-6721-022X
Sara Lindersson 0000-0001-9359-6218
Yoshihide Wada 0000-0003-4770-2539
Marc F P Bierkens 0000-0002-7411-6562
Yadu Pokhrel 0000-0002-1367-216X
Juan Rocha 0000-0003-2322-5459
Giuliano Di Baldassarre 0000-0002-8180-4996
Matti Kummu 0000-0001-5096-0163
Grant Ferguson 0000-0001-8519-1771
Abhijit Mukherjee 0000-0002-0555-0875
Min-Hui Lo 0000-0002-8653-143X
Bridget R Scanlon 0000-0002-1234-4199
Mark S Johnson 0000-0001-5070-7539
Chunmiao Zheng 0000-0001-5839-1305

References

Abbott B W *et al* 2019 Human domination of the global water cycle absent from depictions and perceptions *Nat. Geosci.* **12** 533–40

Aldous A R and Gannett M W 2021 Groundwater, biodiversity, and the role of flow system scale *Ecohydrology* **14** e2342

Andrachuk M and Armitage D 2015 Understanding social-ecological change and transformation through community perceptions of system identity *Ecol. Soc.* **20** 26

Balvanera P, Brauman K A, Cord A F, Drakou E G, Geijzendorffer I R, Karp D S, Martín-López B, Mwampamba T H and Schröter M 2022 Essential ecosystem service variables for monitoring progress towards sustainability *Curr. Opin. Environ. Sustain.* **54** 101152

Barreteau O, Caballero Y, Hamilton S, Jakeman A J and Rinaudo J D 2016 Disentangling the complexity of groundwater dependent social-ecological systems *Integrated Groundwater Management* ed A J Jakeman, O Barreteau, R J Hunt, J D Rinaudo and A Ross (Springer) (https://doi.org/10.1007/978-3-319-23576-9_3)

Berkes F and Folke C 1998 *Linking Social and Ecological Systems: Management Practices and Social Mechanisms for Building Resilience* (Cambridge University Press)

Bierkens M F P 2015 Global hydrology 2015: state, trends, and directions *Water Resour. Res.* **51** 4923–47

Bierkens M F P, Rens van Beek L P H and Wanders N 2024 Gisser-Sánchez revisited: a model of optimal groundwater withdrawal under irrigation including surface–groundwater interaction *J. Hydrol.* **65** 131145

Bierkens M F P and Wada Y 2019 Non-renewable groundwater use and groundwater depletion: a review *Environ. Res. Lett.* **14** 063002

Biggs R, de Vos A, Preiser R, Clements H and Maciejewski K 2021 *The Routledge Handbook of Research Methods for Social-Ecological Systems* ed M Schlüter 1st edn (Routledge) p 526

Binder C R, Hinkel J, Bots P W G and Pahl-Wostl C 2013 Comparison of frameworks for analyzing social-ecological systems *Ecol. Soc.* **18** 1–19

Blomquist W 1994 Changing rules, changing games: evidence from groundwater systems in Southern California *Rules, Games, and Common-Pool Resources* ed E Ostrom, R Gardner and J Walker (University of Michigan Press)

Bloomfield J P, McKenzie A A and Williams A T 2008 An overview of complex behaviour in the groundwater compartment of catchment systems and some implications for modelling and monitoring *Sustainable Hydrology for the 21st Century. 10th BHS National Hydrology Symp.* pp 208–14 (available at: https://nora.nerc.ac.uk/id/eprint/7759)

Blöschl G *et al* 2019 Twenty-three unsolved problems in hydrology (UPH)—a community perspective *Hydrol. Sci. J.* **64** 1141–58

Bojinski S, Verstraete M, Peterson T C, Richter C, Simmons A and Zemp M 2014 The concept of essential climate variables in support of climate research, applications, and policy *Bull. Am. Meteorol. Soc.* **95** 1431–43

Bolognesi T, Gerlak A K and Giuliani G 2018 Explaining and measuring social-ecological pathways: the case of global changes and water security *Sustainability* **10** 4378

Bouchet L, Thoms M C and Parsons M 2019 Groundwater as a social-ecological system: a framework for managing groundwater in Pacific small island developing states *Groundwater Sustain. Dev.* **8** 579–89

Brunner P, Therrien R, Renard P, Simmons C T and Franssen H H 2017 Advances in understanding river-groundwater interactions *Rev. Geophys.* **55** 818–54

Cabello V, Willaarts B, Aguilar M and Del Moral L 2015 River basins as social-ecological systems: linking levels of societal and ecosystem water metabolism in a semiarid watershed *Ecol. Soc.* **20** 20

Carpenter S R *et al* 2015 Plausible futures of a social-ecological system: Yahara watershed, Wisconsin, USA *Ecol. Soc.* **20** 10

Carroll S R *et al* 2020 The CARE principles for indigenous data governance *Data Sci. J.* **19** 1–12
Carroll S R, Herczog E, Hudson M, Russell K and Stall S 2021 Operationalizing the CARE and FAIR principles for indigenous data futures *Sci. Data* **8** 108
Castilla-Rho J C, Rojas R, Andersen M S, Holley C and Mariethoz G 2017 Social tipping points in global groundwater management *Nat. Hum. Behav.* **1** 640–9
Chávez García Silva R *et al* 2024 Multi-decadal groundwater observations reveal surprisingly stable levels in southwestern Europe *Commun. Earth Environ.* **5** 387
Clark M P *et al* 2021 Open science: open data, open models, and open publications? *Water Resour. Res.* **57** e2020WR029480
Cuthbert M O, Gleeson T, Moosdorf N, Befus K M, Schneider A, Hartman J and Lehner B 2019 Global patterns and dynamics of climate-groundwater interactions *Nat. Clim. Change* **9** 137–41
Dalin C, Taniguchi M and Green T R 2019 Unsustainable groundwater use for global food production and related international trade *Glob. Sustain.* **2** e12
Dalin C, Wada Y, Kastner T and Puma M J 2017 Groundwater depletion embedded in international food trade *Nature* **543** 700–4
Dang Q, Lin X and Konar M 2015 Agricultural virtual water flows within the United States *Water Resour. Res.* **51** 973–86
Delwiche K B, Nelson J, Kowalska N, Moore C E, Shirkey G, Tarin T, Cleverly J R and Keenan T F 2024 Charting the future of the FLUXNET network *Bull. Am. Meteorol. Soc.* **105** E466–73
Di Baldassarre G, Viglione A, Carr G, Kuil L, Yan K, Brandimarte L and Blöschl G 2015 Debates—perspectives on socio-hydrology: capturing feedbacks between physical and social processes *Water Resour. Res.* **51** 4770–81
Ellis E C and Ramankutty N 2008 Putting people in the map: anthropogenic biomes of the world *Front. Ecol. Environ.* **6** 439–47
Famiglietti J S 2014 The global groundwater crisis *Nat. Clim. Change* **4** 945–8
Foster S, Chilton J, Nijsten G and Richts A 2013 Groundwater—a global focus on the 'local resource' *Curr. Opin. Environ. Sustain.* **5** 685–95
Gain A K, Giupponi C and Wada Y 2016 Measuring global water security towards sustainable development goals *Environ. Res. Lett.* **11** 124015
Gbondo A M and Michelsen N 2024 Authorship patterns in Africa's groundwater literature: indications for helicopter research? *Earth Sci. Rev.* **255** 104859
GCOS 2024 Essential climate variables (available at: https://gcos.wmo.int/en/essential-climate-variables) (Accessed 8 April 2024)
Giordano M 2009 Global groundwater? Issues and solutions *Annu. Rev. Environ. Resour.* **34** 153–78
Gleeson T *et al* 2020b Illuminating water cycle modifications and Earth system resilience in the Anthropocene *Water Resour. Res.* **56** e2019WR024957
Gleeson T and Cardiff M 2013 The return of groundwater quantity: a mega-scale and interdisciplinary "future of hydrogeology"? *Hydrogeol. J.* **21** 1169–71
Gleeson T, Cuthbert M, Ferguson G and Perrone D 2020a Global groundwater sustainability, resources, and systems in the Anthropocene *Annu. Rev. Earth Planet. Sci.* **48** 431–63
Güntner A *et al* 2024 *Global Gravity-based Groundwater Product (G3P)* (GFZ Data Services) p 1.12
Hall C A, Saia S M, Popp A L, Dogulu N, Schymanski S J, Drost N, van Emmerik T and Hut R 2022 A hydrologist's guide to open science *Hydrol. Earth Syst. Sci.* **26** 647–64
Hazlett M A, Henderson K M, Zeitzer I F and Drew J A 2020 The geography of publishing in the Anthropocene *Conserv. Sci. Pract.* **2** e270
Huggins X, Gleeson T, Castilla-RhoCastilla-Rho J, Holley C, Re V and Famiglietti J S 2023 Groundwater connections and sustainability in social-ecological systems *Groundwater* **61** 463–78
Huggins X, Gleeson T, Villholth K G, Rocha J C and Famiglietti J S 2024 Groundwaterscapes: a global classification and mapping of groundwater's large-scale socioeconomic, ecological, and Earth system functions *Water Resour. Res.* **60** e2023WR036287
IGRAC 2025 *The Global Groundwater Monitoring Network* (IGRAC GGIS) (https://doi.org/10.58154/6Z0Y-DA34)
Jaramillo F *et al* 2024 The potential of hydrogeodesy to address water-related and sustainability challenges *Water Resour. Res.* **60** e2023WR037020
Jasechko S, Seybold H, Perrone D, Fan Y and Kirchner J W 2021 Widespread potential loss of streamflow into underlying aquifers across the USA *Nature* **591** 391–5
Jasechko S, Seybold H, Perrone D, Fan Y, Shamsudduha M, Taylor R G, Fallatah O and Kirchner J W 2024 Rapid groundwater decline and some cases of recovery in aquifers globally *Nature* **625** 715–21
Karlsson S, Srebotnjak T and GonzalesGonzalez P 2007 Understanding the North-South knowledge divide and its implications for policy: a quantitative analysis of the generation of scientific knowledge in the environmental sciences *Environ. Sci. Policy* **10** 668–84
Kim K, Doi Y, Ramankutty N and Iizumi T 2021 A review of global gridded cropping system data products *Environ. Res. Lett.* **16** 093005
Konikow L F and Kendy E 2005 Groundwater depletion: a global problem *Hydrogeol. J.* **13** 317–20
Kreibich H *et al* 2023 Panta Rhei benchmark dataset: socio-hydrological data of paired events of floods and droughts *Earth Syst. Sci. Data* **15** 2009–23
Kuang X *et al* 2024 The changing nature of groundwater in the global water cycle *Science* **383** eadf0630
Lamptey B *et al* 2024 Challenges and ways forward for sustainable weather and climate services in Africa *Nat. Commun.* **15** 2664
Lazurko A, Haider L J, Hertz T, West S and McCarthy D D P 2024 Operationalizing ambiguity in sustainability science: embracing the elephant in the room *Sustain. Sci.* **19** 595–614
Lin C-Y, Miller A, Waqar M and Marston L T 2024 A database of groundwater wells in the United States *Sci. Data* **11** 335
Lin M, Biswas A and Bennett E M 2020 Socio-ecological determinants on spatio-temporal changes of groundwater in the Yellow River Basin, China *Sci. Total Environ.* **731** 138725
Lindersson S, Brandimarte L, Mård J and Di Baldassarre G 2020 A review of freely accessible global datasets for the study of floods, droughts and their interactions with human societies *WIREs Water* **7** e1424
Maas B, Pakeman R J, Godet L, Smith L, Devictor V and Primack R 2021 Women and Global South strikingly underrepresented among top-publishing ecologists *Conserv. Lett.* **14** e12797
McCabe M F *et al* 2017 The future of Earth observation in hydrology *Hydrol. Earth Syst. Sci.* **21** 3879–914
McGinnis M and Ostrom E 2014 Social-ecological system framework: initial changes and continuing challenges *Ecol. Soc.* **19** 1–12
Mehta P, Siebert S, Kummu M, Deng Q, Ali T, Marston L, Xie W and Davis K F 2024 Half of twenty-first century global irrigation expansion has been in water-stressed regions *Nat. Water* **2** 254–61
Meyfroidt P *et al* 2022 Ten facts about land systems for sustainability *Proc. Natl Acad. Sci.* **119** e2109217118
Mills B R 2022 MetBrewer: color palettes inspired by works at the metropolitan museum of art (version 0.2.0) (available at: https://cran.r-project.org/package=MetBrewer)
Miloslavich P *et al* 2018 Essential ocean variables for global sustained observations of biodiversity and ecosystem changes *Glob. Change Biol.* **24** 2416–33

Misstear B, Vargas C R, Lapworth D, Ouedraogo I and Podgorski J 2023 A global perspective on assessing groundwater quality *Hydrogeol. J.* **31** 11–14
Mueller W *et al* 2024 Saltwater intrusion and human health risks for coastal populations under 2050 climate scenarios *Sci. Rep.* **14** 15881
Mukherjee A 2024 *Water Matters: Achieving the Sustainable Development Goals* (Elsevier)
Mukherjee A, Duttagupta S, Chattopadhyay S, Bhanja S N, Bhattacharya A, Chakraborty S, Sarkar S, Ghosh T, Bhattacharya J and Sahu S 2019 Impact of sanitation and socio-economy on groundwater fecal pollution and human health towards achieving sustainable development goals across India from ground-observations and satellite-derived nightlight *Sci. Rep.* **9** 15193
Müller C *et al* 2019 The global gridded crop model intercomparison phase 1 simulation dataset *Sci. Data* **6** 50
Nelson R L and Perrone D 2016 Local groundwater withdrawal permitting laws in the South-Western U.S.: California in comparative context *Groundwater* **54** 747–53
Ostrom E 1990 *Governing the Commons: The Evolution of Institutions for Collective Action* 1st edn (Cambridge University Press) (https://doi.org/10.1017/CBO9780511807763)
Ostrom E 2009 A general framework for analyzing sustainability of social-ecological systems *Science* **325** 419–22
Pacheco-Romero M, Alcaraz-Segura D, Vallejos M and Cabello J 2020 An expert-based reference list of variables for characterizing and monitoring social-ecological systems *Ecol. Soc.* **25** 1
Page M J *et al* 2021 The PRISMA 2020 statement: an updated guideline for reporting systematic reviews *BMJ* **372** 1–9
Pereira H M *et al* 2013 Essential biodiversity variables *Science* **339** 277–8
Preiser R, Biggs R, De Vos A and Folke C 2018 Social-ecological systems as complex adaptive systems: organizing principles for advancing research methods and approaches *Ecol. Soc.* **23** 1–15
R Core Team 2023 R: A language and environment for statistical computing, Version 4.3.1 (available at: https://www.r-project.org/)
Re V 2015 Incorporating the social dimension into hydrogeochemical investigations for rural development: the Bir Al-Nas approach for socio-hydrogeology *Hydrogeol. J.* **23** 1293–304
Reinecke R *et al* 2021 Uncertainty of simulated groundwater recharge at different global warming levels: a global-scale multi-model ensemble study *Hydrol. Earth Syst. Sci.* **25** 787–810
Reinecke R *et al* 2024 Uncertainty in model estimates of global groundwater depth *Environ. Res. Lett.* **19** 114066
Rica M 2017 Understanding groundwater governance through a social ecological system framework—relevance and limits *Advances in Groundwater Governance* ed K G Villholth, E López-Gunn, A Garrido, J A M Gun and K I Conti (CRC Press/Balkema)
Richey A S, Thomas B F, Lo M-H, Reager J T, Famiglietti J S, Voss K, Swenson S and Rodell M 2015 Quantifying renewable groundwater stress with GRACE *Water Resour. Res.* **51** 5217–38
Rohde M M *et al* 2024 Groundwater-dependent ecosystem map exposes global dryland protection needs *Nature* **632** 101–7
Rohde M M, Froend R and Howard J 2017 A global synthesis of managing groundwater dependent ecosystems under sustainable groundwater policy *Groundwater* **55** 293–301
RRI 2015 Who owns the world's land? A global baseline of formally recognized indigenous and community land rights (Rights and Resources Initiative) (https://doi.org/10.53892/NXFO7501)
Saccò M *et al* 2024 Groundwater is a hidden global keystone ecosystem *Glob. Change Biol.* **30** e17066
Scanlon B R *et al* 2023 Global water resources and the role of groundwater in a resilient water future *Nat. Rev. Earth Environ.* **4** 87–101
Schellekens J *et al* 2017 A global water resources ensemble of hydrological models: the eartH2Observe Tier-1 dataset *Earth Syst. Sci. Data* **9** 389–413
Schlüter M, Haider L J, Lade S J, Lindkvist E, Martin R, Orach K, Wijermans N and Folke C 2019 Capturing emergent phenomena in social-ecological systems: an analytical framework *Ecol. Soc.* **24** 11
Scown M W, Flotemersch J E, Spanbauer T L, Eason T, Garmestani A and Chaffin B C 2017 People and water: exploring the social-ecological condition of watersheds of the United States *Elem. Sci. Anth.* **5** 64
Senior C A *et al* 2021 Convection-permitting regional climate change simulations for understanding future climate and informing decision-making in Africa *Bull. Am. Meteorol. Soc.* **102** E1206–E23
Shaminder P and Villholth K G 2017 Governance and management of transboundary aquifers *Advances in Groundwater Governance* ed K G Villholth, E Lopez-Gunn, K Conti, A Garrido and J Van Der Gun (CRC Press) pp 1–22
Siebert S, Burke J, Faures M, FrenkenKrenken K, Hoogeveen J, Doll P and Portmann F T 2010 Groundwater use for irrigation—a global inventory *Hydrol. Earth Syst. Sci.* **14** 1863–80
Siebert S, Kummu M, Porkka M, Döll P, Ramankutty N and Scanlon B R 2015 A global data set of the extent of irrigated land from 1900 to 2005 *Hydrol. Earth Syst. Sci.* **19** 1521–45
Sietz D, Lüdeke M K B and Walther C 2011 Categorisation of typical vulnerability patterns in global drylands *Glob. Environ. Change* **21** 431–40
Song S, Wen H, Wang S, Wu X, Cumming G S and Fu B 2024 Quantifying the effects of institutional shifts on water governance in the Yellow River Basin: a social-ecological system perspective *J. Hydrol.* **629** 130638
Stein L, Mukkavilli S K, Pfitzmann B M, Staar P W J, Ozturk U, Berrospi C, Brunschwiler T and Wagener T 2024 Wealth over Woe: global biases in hydro-hazard research *Earth's Future* **12** e2024EF004590
Sun Q, Miao C, Duan Q, Ashouri H, Sorooshian S and Hsu K 2018 A review of global precipitation data sets: data sources, estimation, and intercomparisons *Rev. Geophys.* **56** 79–107
Taylor R G *et al* 2013 Groundwater and climate change *Nat. Clim. Change* **3** 322–9
Tennekes M *et al* 2018 tmap: thematic maps in R *J. Stat. Softw.* **84** 1–39
Tiwari A D, Pokhrel Y, Kramer D, Akhter T, Tang Q, Liu J, Qi J, Loc H H and Lakshmi V 2023 A synthesis of hydroclimatic, ecological, and socioeconomic data for transdisciplinary research in the Mekong *Sci. Data* **10** 283
Troy T J, Pavao-Zuckerman M and Evans T P 2015 Debates—perspectives on socio-hydrology: socio-hydrologic modeling: tradeoffs, hypothesis testing, and validation *Water Resour. Res.* **51** 4806–14
UNESCO 2021 UNESCO recommendation on open science (United Nations Educational, Scientific and Cultural Organization) (https://doi.org/10.54677/MNMH8546)
Václavík T, Lautenbach S, Kuemmerle T and Seppelt R 2013 Mapping global land system archetypes *Glob. Environ. Change* **23** 1637–47
Varis O, Taka M and Kummu M 2019 The planet's stressed river basins: too much pressure or too little adaptive capacity? *Earth's Future* **7** 1118–35
Villholth K G and Conti K 2018 Groundwater governance: rationale, definition, current state and heuristic framework *Advances in Groundwater Governance* ed K G Villholth, E López-Gunn, A Garrido, J A M Gun and K I Conti (CRC Press/Balkema)
Wada Y, van Beek L P H L, Sperna Weiland F C, Chao B F, Wu Y-H and Bierkens M F P M 2012b Past and future contribution of global groundwater depletion to sea-level rise *Geophys. Res. Lett.* **39** L09402

Wada Y, van Beek L P H L, van Kempen C M, Reckman J W, Vasak S and Bierkens M F 2010 Global depletion of groundwater resources *Geophys. Res. Lett.* **37** L20402
Wada Y, van Beek L P H and Bierkens M F P 2012a Nonsustainable groundwater sustaining irrigation: a global assessment *Water Resour. Res.* **48** W00L06
Wagener T, Gleeson T, Coxon G, Hartmann A, Howden N, Pianosi F, Rahman M, Rosolem R, Stein L and Woods R 2021 On doing hydrology with dragons: realizing the value of perceptual models and knowledge accumulation *WIREs Water* **8** e1550
Walter M, Lovett R, Maher B, Williamson B, Prehn J, Bodkin-Andrews G and Lee V 2021 Indigenous data sovereignty in the era of big data and open data *Aust. J. Soc. Issues* **56** 143–56
Wang Y *et al* 2023 Groundwater quality and health: making the invisible visible *Environ. Sci. Technol.* **57** 5125–36
Wang Y, Köhler P, Braghiere R K, Longo M, Doughty R, Bloom A A and Frankenberg C 2022 GriddingMachine, a database and software for Earth system modeling at global and regional scales *Sci. Data* **9** 258
Warszawski L, Frieler K, Huber V, Piontek F, Serdeczny O and Schewe J 2014 The inter-sectoral impact model intercomparison project (ISI–MIP): project framework *Proc. Natl Acad. Sci. USA* **111** 3228–32
Westerberg I K, Di Baldassarre G, Beven K J, Coxon G and Krueger T 2017 Perceptual models of uncertainty for socio-hydrological systems: a flood risk change example *Hydrol. Sci. J.* **62** 1705–13
Wickham H *et al* 2016 *ggplot2: Elegant graphics for data analysis* 2nd edn (Springer) pp 1–260 (available at: https://ggplot2.tidyverse.org/)
Wilkinson M D *et al* 2016 The FAIR guiding principles for scientific data management and stewardship *Sci. Data* **3** 160018
World Meteorological Organization (WMO) 2022 "Early Warnings for All": the UN global early warning initiative for the implementation of climate adaptation (available at: www.un.org/en/climatechange/early-warnings-for-all) (Accessed 27 January 2025)
World Meteorological Organization (WMO) 2024 State of global water resources 2023 (available at: https://library.wmo.int/idurl/4/69033) (Accessed 25 January 2025)
Wu W-Y W-Y, Lo M-H M-H, Wada Y, Famiglietti J S, Reager J T, Yeh P J-F, Ducharne A and Yang Z-L 2020 Divergent effects of climate change on future groundwater availability in key mid-latitude aquifers *Nat. Commun.* **11** 3710
Xie J, Liu X, Jasechko S, Berghuijs W R, Wang K, Liu C, Reichstein M, Jung M and Koirala S 2024 Majority of global river flow sustained by groundwater *Nat. Geosci.* **17** 770–7
Yang C, Condon L E and Maxwell R M 2025 Unravelling groundwater–stream connections over the continental United States *Nat. Water* **3** 70–79
Yao Y, Tian Y, Andrews C, Li X, Zheng Y and Zheng C 2018 Role of groundwater in the dryland ecohydrological system: a case study of the Heihe River Basin *J. Geophys. Res. Atmos.* **123** 6760–76
Zamrsky D *et al* 2025 Current trends and biases in groundwater modelling using the community-driven groundwater model portal (GroMoPo) *Hydrogeol. J.* **33** 355–66
Zellner M L 2008 Embracing complexity and uncertainty: the potential of agent-based modeling for environmental planning and policy *Plan. Theory Pract.* **9** 437–57
Zipper S C *et al* 2019 Balancing open science and data privacy in the water sciences *Water Resour. Res.* **55** 5202–11
Zipper S *et al* 2023 GroMoPo: a groundwater model portal for findable, accessible, interoperable, and reusable (FAIR) modeling *Groundwater* **61** 764–7
Zwarteveen M *et al* 2021 Transformations to groundwater sustainability: from individuals and pumps to communities and aquifers *Curr. Opin. Environ. Sustain.* **49** 88–97