Global Dam Watch (GDW) Database

A harmonized and curated database of river barriers and reservoirs worldwide

Technical Documentation – version 1.0

prepared by Bernhard Lehner (bernhard.lehner@mcgill.ca)
on behalf of Global Dam Watch (www.globaldamwatch.org)

July 2024

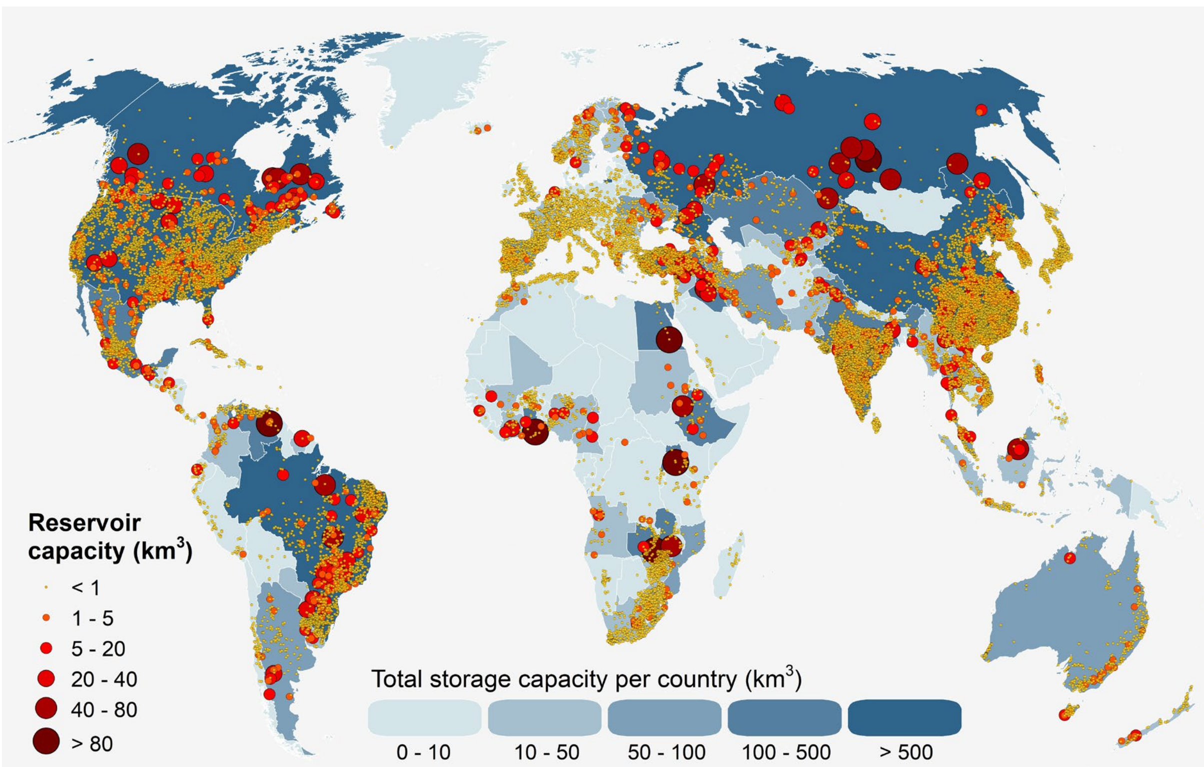

Figure 1: Global distribution of river barriers and reservoirs in GDW version 1.0 database.

Note:
This document refers to version 1.0 of the Global Dam Watch (GDW) database. The data are freely available from the Global Dam Watch website at www.globaldamwatch.org and from the figshare repository at https://doi.org/10.6084/m9.figshare.25988293.

1. Overview and background

Despite established recognition of the many critical environmental and social tradeoffs associated with dams, other instream barriers, and their reservoirs, global datasets describing their characteristics and geographical distribution have been largely incomplete or are biased towards particular regions or applications. There are likely millions of dams, river barriers, and reservoirs worldwide (Lehner et al. 2011), but despite multiple efforts by individual groups, only a small proportion of them have been mapped today.

The development of the Global Dam Watch (GDW) database has been coordinated by the Global Dam Watch consortium (see www.globaldamwatch.org) which was initiated by several academic institutions and NGOs that work together to fill critical gaps of global dam and reservoir information. A particular goal of the Global Dam Watch initiative is to advance recent efforts to develop a single, harmonized and curated global data product of dams, other instream barriers, and reservoirs for large-scale analyses: the GDW database (Mulligan et al. 2021). For this task, existing data repositories are compiled, cleaned, and curated, and new data are being collected using a variety of methods, from citizen science to remote sensing and machine learning. Results of the different approaches are harmonized to create consistent, high quality river barrier and reservoir information at the global scale. The GDW database aims to include all types of anthropogenic instream barriers, though initial mapping efforts prioritize major dams that form reservoirs, as well as run-of-river barriers on larger rivers, for which more information is available.

The current version of the GDW database (version 1.0) contains 41,145 barrier and dam locations and 35,295 associated reservoir polygons (Figure 1). The database is freely available for download at www.globaldamwatch.org. This Technical Documentation provides an overview and explains the technical specifications of the database. The development and characteristics of GDW v1.0 are fully described in Lehner et al. (2024) and should be cited as:

Lehner, B., Beames, P., Mulligan, M., Zarfl, C., De Felice, L., van Soesbergen, A., Thieme, M., Garcia de Leaniz, C., Anand, M., Belletti, B., Brauman, K.A., Januchowski-Hartley, S.R., Lyon, K., Mandle, L., Mazany-Wright, N., Messager, M.L., Pavelsky, T., Pekel, J.-F., Wang, J., Wen, Q., Wishart, M., Xing, T., Yang, X., Higgins, J. (2024): The Global Dam Watch database of river barrier and reservoir information for large-scale applications. Scientific Data. [please insert journal volume and DOI once released]

2. Methods

The details of the GDW v1.0 database development are described in Lehner et al. (2024). During various steps of data consolidation and harmonization, extensive manual inspections were carried out, and a variety of Geographic Information System (GIS) techniques were applied to detect potential errors or issues in the provided data, including inconsistencies in spatial location, attribute information, or potential duplicate records. The locations of all barriers, dams and reservoirs were verified through manual or (supervised) automated processes, and the data records were updated and/or newly georeferenced as needed. This manual curation process was guided by a variety of online digital mapping resources, including Google Earth and ESRI Basemaps. The development of the GDW database was coordinated by the Global Dam Watch consortium and was executed in partnership and collaboration between members of the following institutions and organizations: McGill University, Montreal, Canada; King's College London, UK; University of Tübingen, Germany; the European Commission's Joint Research Center, Ispra, Italy; the University of North Carolina, Chapel Hill, USA; Swansea University, UK; and World Wildlife Fund, Washington DC, USA.

2.1 Main data sources

The development of version 1.0 of the GDW database is primarily aimed at compiling available global barrier, dam, and reservoir information; harmonizing and curating it through both (supervised) automated and manual cross-validation, error checking, and identification of duplicate records, attribute conflicts, or mismatches; and augmenting missing information from a multitude of sources or statistical approaches. Table 1 describes the main input datasets used in this process. While the extent of all these data repositories is fully global, they show different characteristics regarding their content, comprehensiveness, and the type of attributes they provide. Differences are mostly due to the objectives of each dataset and the underpinning sources used to assemble them. For example, many of the sources for the GRanD database used a height threshold of 15 m for dams in their original collections, introducing a bias in the initial selection towards higher and larger dams.

Table 1: *Main data sources used in the development of the GDW database, their characteristics, and the number of included records. It should be noted that these collections, in turn, used underpinning information from a much wider range of sources which can be found in their respective reference papers.*

Dataset	***Reference***	***Data characteristics or main purpose in creation of GDW database***	***Contributed objects and attributes***	***Number of contributed records****
GOODD (GlObal geOreferenced Database of Dams)	Mulligan et al. 2020	Locations of medium to large dams that are visible on satellite imagery (Google Earth); dam ≥150 m long and reservoir ≥500 m long; manually digitized	Barrier points	25,931
GRanD (Global Reservoir and Dam database)	Lehner et al. 2011	Large dams and reservoirs (≥0.1 km^3); compiled from freely available data, peer-reviewed and grey literature, internet; manual inspection and validation of all records; extensive attribute information	Barrier points and reservoir polygons; multiple attributes incl. name, year, height, purpose, reservoir volume	7,424
FHReD (Future Hydropower Reservoirs and Dams database)	Zarfl et al. 2015	Hydropower dams ≥1 MW; compiled from freely available data, peer-reviewed and grey literature, internet; manual inspection and validation of all records; original dataset focused on planned projects, from which those completed by 2022 were selected	Barrier points; hydropower capacity, year of construction	205
JRC-GSW (Global Surface Water Explorer of European Commission's Joint Research Centre)	Pekel et al. 2016	Surface water extents mapped at 30 m grid resolution from Landsat imagery; automatic extraction of new reservoirs that appeared after 1984	Reservoir polygons; years of construction inferred from time series of satellite imagery	1,451 new reservoir polygons (and 14,015 polygons for existing barriers)
GROD (Global River Obstruction Database)	Yang et al. 2022	Instream barriers (incl. dams, locks, and other barrier types) on rivers wider than 30 m, mapped through manual detection from remote sensing imagery	Barrier points; barrier type	6,113
HydroLAKES	Messager et al. 2016	Polygons of all lakes globally with a surface area ≥10 ha; polygons were used as reservoir outlines if they were associated with a barrier from GOODD, FHReD or GROD	Reservoir polygons and barrier points (lake outlets)	No new records (but source of 13,854 reservoir polygons)
HydroSHEDS and RiverATLAS	Lehner et al. 2008; Linke et al. 2019	Digital river network to which the barrier/dam locations were co-registered; after co-registration, some hydrometric attributes were derived	Catchment area, long-term mean discharge, degree of regulation	No new records (but source of attributes for all records)

* The original number of available records per dataset may be higher; it is reduced here due to removal of duplicates.

The five foundational source datasets from which the first version of the GDW database was created are: 1) GOODD (GlObal geOreferenced Database of Dams; Mulligan et al. 2020); 2) GRanD (Global Reservoir and Dam database, version 1.4; Lehner et al. 2011); 3) FHReD (Future Hydropower Reservoirs and Dams database; Zarfl et al. 2015); 4) JRC-GSW (Global Surface Water Explorer of the European Commission's Joint Research Centre; Pekel et al. 2016); and 5) GROD (Global River Obstruction Database; Yang et al. 2022). All barriers and dams were geospatially referenced as point coordinates and co-registered to the global river network of HydroSHEDS (Lehner et al. 2008). Where possible, the barrier/dam records were associated with reservoir polygons; for this purpose, reservoir outlines were either sourced from the global HydroLAKES database (Messager et al. 2016) or derived from the surface water extent maps of the JRC-GSW database.

While the GDW database aims to include all types of anthropogenic instream barriers, mapping efforts for version 1.0 prioritized major dams that form larger reservoirs, as well as instream barriers on larger rivers, for which more information was available. This focus on 'larger' structures was already inherent in the source datasets used in the compilation of the GDW database. For example, the intent of the GRanD database was to include all reservoirs with a storage capacity of more than 0.1 km^3; the GOODD database mapped medium to large dams visible in publicly accessible remote sensing imagery; FHReD focused exclusively on proposed hydropower dams with a hydropower capacity exceeding 1 MW; and GROD mapped river barriers for rivers wider than 30 meters.

2.2 Creation of corresponding barrier (point) and reservoir (polygon) objects

The majority of source records (i.e., those from the GOODD, FHReD, and GROD datasets) provided only the point locations of barriers and dams, whereas the GRanD database also included polygons of the impounded reservoirs and the JRC-GSW data provided only polygons, without explicit dam information. As a first consolidation step, additional reservoir polygons were created for all barrier or dam locations that could be associated with a storage reservoir. Many of these polygons were sourced from the HydroLAKES dataset (Messager et al. 2016): corresponding polygons were either extracted through an automated 'spatial join' procedure (i.e., identified by barrier points that fell inside or were within 1 km of an existing lake polygon from HydroLAKES), or by manual inspections of candidate polygons that were in close vicinity (1-5 km) of barrier or dam locations. In addition, new polygons were created by converting rasterized open water extents from the JRC-GSW dataset into polygons (see section 2.3 below for details). The new JRC-GSW polygons were manually inspected for correctness and were modified as needed. Finally, in few instances entirely new polygons were digitized. It should be noted that reservoir outlines are typically subject to strong seasonal fluctuations; and as many polygons included in the GDW database are originally depicted from remote sensing imagery (i.e., a snapshot in time) they may reflect a low-fill or dry-season state with significantly smaller than maximum area.

In a second consolidation step, each reservoir was associated with one representative dam location. For records derived from the GRanD database, this information already existed in the original source data. For reservoir polygons added from the HydroLAKES dataset, the existing outlet points of the HydroLAKES polygons were used as a proxy for the associated dam locations. For newly created polygons (i.e., mostly those from the JRC-GSW data), the barrier locations were derived as the pixel with the highest upstream flow accumulation within the reservoir polygon according to the HydroSHEDS drainage maps (Lehner et al. 2008). All barrier points were placed inside the intersection between the respective reservoir polygon and the selected pixel. Some exceptions and corrections were applied during manual inspections.

As a result of this processing workflow, each record in the GDW database—as identified by a unique ID—typically represents a paired 'barrier-and-reservoir object' which is defined by both a point location and a polygon outline (see also section 3.1 on data formats). The point represents the location of the barrier or dam, or the 'main' dam in case of multiple barriers forming a single reservoir (these latter cases are further described in columns 'Multi_dams' and 'Comments' in the attribute table). Furthermore, barrier objects can also be defined by a point only, representing an independent barrier or dam without a 'traditional' reservoir, including run-of-river hydropower stations, navigation locks, diversion barrages, check dams that only briefly create storage reservoirs during flood events, weirs and other instream control barriers, or dams under construction that do not yet have a filled reservoir.

2.3 Procedures for creating new reservoir polygons from JRC-GSW data

For the creation of the GDW database, many new reservoir polygons were delineated from the surface water maps of the JRC-GSW data product, which were produced from Landsat imagery at 30 m resolution (Pekel et al. 2016). For the creation of the GDW v1.0 database, the JRC-GSW maps showing 'maximum surface water extent' were used for the period 1984-2022. The gridded data were first modified with boundary cleaning filters to consolidate connected water surfaces and to slightly smooth the shorelines and were then converted to polygons. After reservoir shorelines were created, the polygons were manually inspected and, if necessary, corrected by consulting remote sensing imagery and any auxiliary documents pertaining to the reservoir. In particular, adjustments were made, mostly by visual image interpretation, to isolate the reservoir from inflowing rivers, or to merge multiple pools which were falsely separated by a bridge or due to a narrow channel. In some instances where a reservoir was not visible in the JRC-GSW data as it was not yet filled in the year of data provision, or obscured by persistent clouds, reservoir polygons were manually delineated based on ESRI Basemaps and/or other georeferenced imagery. Some remaining dam locations had no visible reservoir in any available imagery; they were annotated as not yet filled ("no polygon") in the point version of the GDW database, and no associated reservoir record exists in the polygon version.

2.4 Identification and removal of duplicates

Linking the original records of all source datasets to the same polygon features introduced a clear relationship between reservoirs and their associated barrier(s), which supported the identification and elimination of duplicate barriers. If dam or barrier points from multiple source datasets were associated with the same reservoir polygon, they were considered duplicates and only one consolidated record was kept in the GDW database.

For barrier and dam locations without reservoirs, duplicates were harder to detect. In iterative, semi-automated detection procedures, point locations were assigned the distance to their nearest neighboring point. All points closer than 2 km from another point or reservoir polygon were flagged and manually inspected as to whether they resembled the same object.

2.5 Co-registration to a global river network

In order to assign each barrier or dam to a representative location on a river, they were co-registered to the global digital river network of the HydroSHEDS database (Lehner et al. 2008). All records represented by a barrier point only (i.e., without a reservoir) were manually allocated to the nearest 'topologically correct' pixel in HydroSHEDS (i.e., to the correct river mainstem or tributary). This process was guided by remote sensing imagery (mostly Google Earth and ESRI Basemaps). For records with a reservoir polygon, the reservoir's outlet point was used as a proxy for its barrier location (see section 2.2 above), which by default is located inside the raster cell that represents the main river draining the reservoir.

It should be noted that although visual inspections showed good spatial correspondence between the barrier points, reservoir polygons, and the river network of HydroSHEDS, spatial offsets and uncertainties in the range of 500 m are inherent in the river delineations due to the applied raster cell resolution of 15 arc-seconds. Therefore, the representative barrier location on the river network is only an approximation of the true dam location. For some records, both the original dam location and the representative location on the river network were recorded (see Table 2).

2.6 Derivation of general barrier/dam and reservoir attribute information

A broad range of attribute information for dams and reservoirs was available in the GRanD database. Other source datasets offered only specific information, such as hydropower capacity in the FHReD dataset. Where available, reported information from these sources was transferred to the GDW database. Additional attributes were inserted from alternative sources, including regional datasets. E.g., available dam and reservoir information was added from the US National Inventory of Dams (NID; USACE 2021) through a spatial join to the nearest reservoir polygon (up to a distance limit of 500 m).

Furthermore, the linkage of the GDW records with the multiple information layers of the related RiverATLAS database (Linke et al. 2019) allowed for the derivation of additional attributes, in particular catchment area and long-term mean discharge. The discharge values provided by HydroATLAS are based on downscaled runoff estimates from the global hydrological model WaterGAP (Müller Schmied et al. 2021) for the period 1971-2000 and were also used to calculate the 'Degree of Regulation (DOR)' index for every reservoir (see Table 2).

2.7 Estimating missing reservoir volumes

During the development of the GDW database, two regression models were derived and applied to complete missing reservoir volumes, following the approach by Lehner et al. (2011):

$$V = 0.553\ (A \cdot h)^{0.941} \qquad \text{(Eq. 1)}$$

$$V = 15.662\ A^{1.059} \qquad \text{(Eq. 2)}$$

where V = storage volume of the reservoir in 10^6 m^3; A = surface area of the reservoir in km^2; and h = dam height in m.

Both equations were determined through a bias-corrected power law regression analysis of 7,348 reservoirs worldwide contained in the GDW v1.0 database which were selected based on data reliability using the following criteria: each record showed a reported reservoir volume, a reported dam height, and a calculated surface area from the associated reservoir polygon; the calculated mean depth of each reservoir (reported volume divided by polygon area) was less than the reported dam height and more than 1 m (to exclude potential lake control structures); and the quality of the record was reported as 'Fair' or better. Four additional records in GDW v1.0 matched these requirements but were dismissed as clear outliers after inspecting the regression scatter plots. Equation 1 was used to estimate the missing storage volumes of 89 reservoirs for which both area and dam height were available (R^2 = 0.95 for reservoirs used in the determination of the equation's parameter settings); Equation 2 was used to estimate the missing storage volumes of 25,504 reservoirs for which only the surface area was available ($R^2 = 0.82$)

It should be noted that Equations 1 and 2 were derived by relating reported storage capacities to measured polygon areas. As the polygons in many cases depict a status below full capacity, the equations may not be appropriate to estimate capacities from maximum reported areas. In instances where natural lakes are regulated by dams, such as Africa's Lake Victoria, reported reservoir storage volumes were used; if absent, volumes were estimated from reported regulated lake depth, or by assuming a 1 m depth otherwise (such estimates were only made for 72 records).

2.8 Estimating the filling year for reservoirs built after 1984

For all records in the final GDW database that did not have a reported year of construction but could be associated with a reservoir polygon (n = 6,931), an estimate of the filling year was made in a two-step approach. First, a 'candidate' year was estimated from the JRC-GSW time series data through a heuristic statistical analysis to detect abrupt changes within the reservoir polygon from a non-water to a water surface. Second, each of these candidate years was verified (and corrected if needed) through manual inspection using timelapse remote sensing imagery built from the Landsat archive on Google Earth Engine (see https://earthengine.google.com/timelapse/). Reservoirs that were already filled before the first Landsat imagery was available in 1984 were flagged as 'before 1985'.

While distinct changes in the timelapse sequences were observed for many records, some cases were ambiguous, either due to unclear imagery (e.g., blurred or cloud-covered scenes) or if the filling occurred close to the year 1984 (as a first visible detection of a full reservoir, say, in 1986 could also represent a reservoir that was built much longer ago, yet was empty in 1984 and 1985 due to climate fluctuations or management decisions). In all ambiguous cases (n = 839) filling years were therefore recorded as 'before *YEAR*' where *YEAR* refers to the first clear image of the reservoir. In a test against 111 reservoirs in the US for which years were provided in the US NID dataset, the independently made timelapse estimates were within ±5 years from the reported year for 102 records (92% of cases, including those that were correctly predicted as 'before 1985'), within ±3 years for 98 records (88% of cases), and within ±1 years for 91 records (82% of cases). This demonstrates a good overall reliability of this estimation method.

2.9 Uncertainties, 'quality' flag, and validation

To assess data quality, attribute information for each barrier and reservoir was compiled and cross-referenced using multiple sources to verify veracity and identify conflicts. Links to source materials were included in the respective record for reference where available. Verification efforts were performed using a combination of published information and web-based satellite and reference maps. As a result, some data errors were detected and corrected, or data gaps were filled during the consolidation and curation procedures, e.g., by consulting and adding independent sources of information, or by applying statistical approaches. To indicate an overall estimate of reliability, a generic quality indicator (i.e., *Verified*, *Good*, *Fair*, *Poor*, or *Unreliable*; see Table 2) was assigned to each record by the data editors. Although subjective, this indicator allows identification of records where obvious inconsistencies, uncertainties, or data gaps remain.

Despite these curation efforts, each barrier, dam, or reservoir included in the GDW database is affected by uncertainties in its respective source dataset(s). These uncertainties can relate to the location of the barrier or reservoir, or to its associated attribute information. For example, potential inconsistencies in the GRanD database include typos and order-of-magnitude errors, such as mistyped volumes by a factor of 1000; or unit mismatches (e.g., feet vs. meters). Also, in many instances the dam name is different from the reservoir name, such as Lake Mead, the largest reservoir of the US, being impounded by the Hoover Dam, making attribute associations more difficult. Another uncertainty is caused by the lack of one-to-one relationships between barriers and reservoirs: some dams, such as barrages, diversions, or run-of-river hydropower stations, may not form reservoirs; some reservoirs may have multiple dams (e.g., main and saddle dams); and some reservoirs have no dam at all, such as water stored in natural or artificial depressions. These ambiguities compound the importance of knowing from which source dataset the record was derived; this information is available as part of the GDW attributes (see Table 2).

For additional validation and improvement purposes, attribute information listed by the International Commission on Large Dams (ICOLD) in their World Register of Dams (WRD; ICOLD 1998-2022) was consulted for some dams. Similarly, the recent publication of the GeoDAR dataset (Georeferenced global Dams And Reservoirs; Wang et al. 2022) offered the opportunity to correct some erroneous entries (~90 errors of original GRanD records were flagged through comparison with GeoDAR and subsequently corrected in the GDW database).

3. Data specifications

3.1 File and data formats

The GDW database consists of two separate GIS layers:

- 'GDW_barriers_v1_0' is a point layer containing all estimated barrier locations and their attribute information
- 'GDW_reservoirs_v1_0' is a polygon layer containing all corresponding reservoir outlines and their attribute information

Each barrier point lies within its corresponding reservoir polygon, thus the features and attributes of both layers can be spatially joined based on their location. Additionally, both attribute tables carry the same unique identification number (column 'GDW_ID') for each paired barrier-and-reservoir object. Version 1.0 of the GDW database contains 41,145 barrier points and 35,295 associated reservoir polygons. That is, 5,850 barrier locations have no polygon, including navigation locks, diversion barrages, check dams that create storage only during flood events, weirs and other instream control barriers, or dams under construction that do not yet have a filled reservoir.

Both the point and polygon layer of the GDW database are provided in ESRI© Geodatabase and Shapefile formats. Each shapefile consists of six core files (.cpg, .dbf, .sbn, .sbx, .shp, .shx); and projection information is provided in an ASCII text file (.prj). The data are unprojected using a Geographic Coordinate System with the horizontal datum of the World Geodetic System 1984 (GCS_WGS_1984). The GDW database includes a copy of the GDW Technical Documentation. NOTE: For users without GIS software, the attribute table of the barrier layer has also been included as a stand-alone text file (.txt) in comma delimited UTF-8 format as part of the Shapefile package. This text file contains all GDW attribute information, and the barrier locations can be plotted using the provided x/y-coordinates.

3.2 Attribute table of GDW records

Due to the high variability in the information pertaining to the primary data sources, decisions had to be made regarding which attributes to include in the construction of the GDW database. These decisions were largely driven by requests from users working in different disciplines interested in the application of the GDW database, including hydrology, geomorphology, ecology, biogeochemistry, biodiversity conservation, and water resources management. Depending on data availability, some attribute fields are fully populated, while others remain incomplete. A full list of available attribute columns and their definition is provided in Table 2.

Table 2: *Attributes provided in the point layer (GDW_barriers) and in the polygon layer (GDW_reservoirs) of the GDW database. Note that the 'number of occurrences' refers to the point layer (41,145 dams) and will be lower for the polygon layer (35,295 polygons). The expressions 'dam' and 'barrier' are interchangeable in this table.*

Column title	*Description*	*Number of occurrences*
GDW_ID	Unique ID for each barrier and associated reservoir; IDs correspond between barrier (point) and reservoir (polygon) layers of the GDW database	41,145
Res_name	Name of reservoir or lake (i.e., impounded waterbody)	2,098
Dam_name	Name of dam/barrier structure	10,071
Alt_name	Alternative name of reservoir or dam/barrier (including different spelling or different language)	1,806
Dam_type	Indicates the type of the dam/barrier: *'Dam' (note that this is the default value assigned to all barriers unless another type is known)* *'Low Permeable Dam' (as defined in the GROD database; Yang et al. 2022)* *'Lock' (as defined in the GROD database; Yang et al. 2022)* *'Lake Control Dam' (see also column 'Lake_ctrl' below)*	41,145 *38,910* *886* *1,152* *197*
Lake_ctrl	Indicates whether a reservoir has been built at the location of an existing natural lake using a lake control structure; currently this column only contains limited entries: *'Yes' = lake control structure raises original lake level* *'Maybe' = not verified, but data seem to indicate a lake control structure* *'Enlarged' = lake control structure enlarged the original lake surface area*	209 *172* *33* *4*
River	Name of impounded river	9,501
Alt_river	Alternative name of impounded river (including different spelling or different language)	714
Main_basin	Name of main basin	2,738
Sub_basin	Name of sub-basin	721
Country	Name of country (as defined in the GADM database version 4.1; https://gadm.org)	41,145
Sec_cntry	Secondary country (indicating international dams or reservoirs that lie within or are associated with multiple countries)	202
Admin_unit	Name of administrative unit (as defined in the GADM database version 4.1; https://gadm.org)	41,145
Sec_admin	Secondary administrative unit (indicating dams or reservoirs that lie within or are associated with multiple administrative units; but may also include different spelling or different language)	4,866
Near_city	Name of nearest city	6,370
Alt_city	Alternative name of nearest city (including different spelling or different language)	302
Year_dam	Year in which the dam/barrier was built (not further specified: year of construction; year of completion; year of commissioning; year of refurbishment/update; etc.); either reported or estimated (see also columns 'Pre_year' and 'Year_src')	15,230
Pre_year	Estimated year before which the barrier was built (e.g., 1985 in this column means 'before 1985') as the reservoir was detectable on time-lapse remote sensing imagery thereafter but not before, either due to lack of imagery or unclear imagery; note that the earliest time-lapse imagery used in this estimation was from 1984	2,518

Column title	***Description***	***Number of occurrences***
Year_src	Source of information for 'Year_dam' or 'Pre_year': *'Estimated' (estimated by analyzing time-lapse data of remote sensing imagery; if a clear filling year could be detected, it is recorded in column 'Year_dam'; if filling year is ambiguous, e.g., due to blurry imagery, it is recorded in column 'Pre_year')* *'GRanD' (reported in GRanD database; Lehner et al. 2011)* *'JRC-GSW' (derived through AI-supported auto-detection from JRC-GSW data; Pekel et al. 2016)* *'NID' (reported in NID database; USACE 2021)* *'Other' (reported in other sources)*	17,749 *6,931* *7,071* *1,431* *2,305* *11*
Alt_year	Alternative year of construction (not further specified: may indicate a multi-year construction phase, an update, or a secondary dam construction)	805
Rem_year	Year in which the dam/barrier was removed, replaced, subsumed, or destroyed; see also column 'Timeline' below	10
Timeline	Indicates whether the status of a dam/barrier has changed or is expected to change over time: *'Destroyed' (dam got destroyed or failed)* *'Modified' (dam was modified from an earlier status, e.g., raised, expanded, refurbished, but the earlier status is not individually recorded)* *'Planned' (dam is planned to be built in the future)* *'Removed' (dam record is retained but the dam itself has been removed and not replaced)* *'Replaced' (dam record is retained in dataset but the dam itself has been replaced; the new dam is recorded as a new point)* *'Subsumed' (dam record is retained in dataset but the dam itself was subsumed by larger infrastructure constructed further downstream; the new dam and reservoir are recorded as a new point and polygon)* *'Under construction' (dam is currently under construction)*	70 *2* *53* *3* *5* *3* *2* *2*
Year_txt	Summary of year information in text format	41,145
Dam_hgt_m	Height of dam/barrier in meters	9,311
Alt_hgt_m	Alternative height of dam/barrier (may indicate an update or secondary dam construction)	366
Dam_len_m	Length of dam/barrier in meters	8,276
Alt_len_m	Alternative length of dam/barrier (may indicate an update or secondary dam construction)	208
Area_skm	Representative surface area of reservoir in square kilometers; consolidated from other 'Area' columns in the following order of priority: 'Area_poly' over 'Area_rep' over 'Area_max' over 'Area_min'; some exceptions apply if value in 'Area_poly' column seems unreliable; see also additional notes below the table	35,321
Area_poly	Surface area of associated reservoir polygon in square kilometers	35,295
Area_rep	Most reliable reported surface area of reservoir in square kilometers	7,444
Area_max	Maximum value of other reported surface areas in square kilometers	158
Area_min	Minimum value of other reported surface areas in square kilometers	289
Cap_mcm	Representative maximum storage capacity of reservoir in million cubic meters; consolidated from other 'Cap' columns in the following order of priority: 'Cap_max' over 'Cap_rep' over 'Cap_min'; some exceptions apply if value in 'Cap_max' column seems unreliable or rounded; if no capacity was reported, it was estimated using statistical approaches (see section 2.7); see also additional notes below the table	35,334
Cap_max	Reported 'maximum storage capacity' in million cubic meters; see also notes below the table	4,403
Cap_rep	Reported 'storage capacity' in million cubic meters; value may refer to different types of storage capacity; see also notes below the table	9,044
Cap_min	Minimum value of other reported storage capacities in million cubic meters	1,176
Depth_m	Average depth of reservoir in meters; calculated as ratio between storage capacity ('Cap_mcm') and surface area ('Area_skm'); values that are somewhat higher than the dam height ('Dam_hgt_m') may still be reasonable, e.g. if the storage capacity refers to the maximum volume yet the reservoir polygon represents a low-fill status; values capped at 1000 indicate exceedingly high values which may be due to inconsistencies in the data	35,321

Column title	Description	Number of occurrences
Dis_avg_ls	Long-term (1971-2000) average discharge at barrier location in liters per second; value derived from HydroSHEDS flow routing scheme combined with downscaled WaterGAP runoff estimates (Müller Schmied et al. 2021) at 15-sec resolution at point location of barrier (Linke et al. 2019)	41,134
Dor_pc	Degree of Regulation (DOR) in percent; equivalent to "residence time" of water in the reservoir; calculated as ratio between storage capacity ('Cap_mcm') and total annual flow (derived from 'Dis_avg_ls'); values capped at 10,000 indicate exceedingly high values, which may be due to inconsistencies in the data and/or incorrect allocation to the river network and the associated discharges	35,168
Elev_masl	Elevation of reservoir surface in meters above sea level; value derived from EarthEnv-DEM90 dataset (Robinson et al. 2014) at 15-sec resolution as minimum within reservoir polygon or at point location of barrier, respectively	41,134
Catch_skm	Area of upstream catchment draining into the reservoir in square kilometers; value derived from HydroSHEDS at 15-sec resolution at point location of barrier (Linke et al. 2019)	41,134
Catch_rep	Reported area of upstream catchment draining into reservoir in square kilometers	4,007
Power_mw	Hydropower capacity in MW	242
Data_info	Supporting information on certain data issues: *'Capacity from statistics' = capacity derived from Eq. 1 or Eq. 2* *'Capacity estimated' = capacity estimated from other available information (including the assumption of a regulation depth of ~1 m for controlled lakes)* *'NID data' = capacity and/or other geometric information converted from US NID*	27,977 *25,528* *80* *2,369*
Use_irri	Used for irrigation ('Main'; 'Major'; 'Sec' = Secondary use; or 'Multi' if multiple uses exist without a ranking); see also additional notes below the table	2,669
Use_elec	Used for hydroelectricity production ('Main'; 'Major'; 'Sec'; or 'Multi')	3,065
Use_supp	Used for water supply ('Main'; 'Major'; 'Sec'; or 'Multi')	2,286
Use_fcon	Used for flood control ('Main'; 'Major'; 'Sec'; or 'Multi')	2,030
Use_recr	Used for recreation ('Main'; 'Major'; 'Sec'; or 'Multi')	2,105
Use_navi	Used for navigation ('Main'; 'Major'; 'Sec'; or 'Multi')	322
Use_fish	Used for fisheries ('Main'; 'Major'; 'Sec'; or 'Multi')	359
Use_pcon	Used for pollution control ('Main'; 'Major'; 'Sec'; or 'Multi')	106
Use_live	Used for livestock water supply ('Main'; 'Major'; 'Sec'; or 'Multi')	49
Use_othr	Used for other purposes ('Main'; 'Major'; 'Sec'; or 'Multi'); other purposes may include new or a mix of the above purposes	800
Main_use	Main purpose of reservoir: *Irrigation; Hydroelectricity; Water supply; Flood control; Recreation; Navigation; Fisheries; Pollution control; Livestock; Other;* or *Multipurpose* (if multiple uses exist without a ranking); see also additional notes below the table	8,435
Multi_dams	Indicates whether there is more than one dam/barrier associated with this reservoir (e.g., main and saddle dam); if 'Yes', then columns 'Alt_year', 'Alt_hgt_m', and 'Alt_len_m' refer to the secondary dam	225
Comments	Comments	964
Url	URL of related website	1,229
Quality	Quality index: *1: Verified (location and all attributes have been fully verified)* *2: Good (location and data seem good but not all attributes have been verified)* *3: Fair (some data discrepancies; missing data; or uncertainties)* *4: Poor (significant data discrepancies of various kinds that indicate errors)* *5: Unreliable (severe data discrepancies without reasonable explanation)*	41,145 *31* *8,118* *32,456* *470* *70*
Editor	Final data editor: *'McGill' = McGill University (BL = B. Lehner; PB = P. Beames; MA = M. Anand; TX = T. Xing)* *'UNH' = University of New Hampshire (as part of GRanD database; Lehner et al. 2011)*	41,145 *39,260* *1,885*

Column title	Description	Number of occurrences
Long_riv	Longitude of the point location of the dam/barrier in decimal degrees after it was associated with a river segment of HydroSHEDS; i.e., the point location is only an approximation of the actual dam/barrier location; this is the location of the point as provided in the GIS layer	41,145
Lat_riv	Latitude of the point location of the dam/barrier in decimal degrees after it was associated with a river segment of HydroSHEDS; see associated 'Long_riv' column for more details	41,145
Long_dam	Longitude of the actual point location of the dam/barrier in decimal degrees; i.e., this represents the actual location of the dam/barrier before it was associated with a river segment of HydroSHEDS; this information is not available for records that were originally mapped to the river network or reservoir polygon without detailed detection of the true dam/barrier location	6,113
Lat_dam	Latitude of the actual point location of the dam/barrier in decimal degrees; see associated 'Long_dam' column for more details	6,113
Orig_src	Original dataset from which the dam/barrier or reservoir was derived: *'FHReD' = Future Hydropower Reservoirs and Dams database (Zarfl et al. 2015)* *'GOODD' = GlObal geOreferenced Database of Dams (Mulligan et al. 2020)* *'GOODD-NID' = GOODD with attribute information from NID (USACE 2021)* *'GRanD' = Global Reservoir and Dam database v1.4 (Lehner et al. 2011)* *'GROD' = Global River Obstruction Database (Yang et al. 2022)* *'GROD-NID' = GROD with attribute information from NID (USACE 2021)* *'JRC-GSW' = Global Surface Water Explorer of the European Commission's Joint Research Centre (Pekel et al. 2016)* *'JRC-NID' = JRC-GSW with attribute information from NID (USACE 2021)* *'Other' = other data source, including original mapping by McGill University*	41,145 *205* *23,633* *2,298* *7,424* *6,060* *53* *1,426* *25* *21*
Poly_src	Original source of reservoir polygon: *'CanVec' = Canadian hydrographic dataset (Natural Resources Canada 2013)* *'ECRINS' = European Catchments and Rivers Network System (EEA 2012)* *'GLWD' = Global Lakes and Wetlands Database (Lehner & Döll 2004)* *'JRC-GSW' = polygon digitized from the gridded dataset of the Global Surface Water Explorer of the European Commission's Joint Research Centre (Pekel et al. 2016)* *'JRC-GSW-mod' = initial polygon digitized from JRC Global Surface Water Explorer data and then modified by McGill University* *'McGill' = polygon digitized from scratch or majorly modified by McGill University* *'SWBD' = SRTM Water Body Database (Slater et al. 2006)* *'UY' = polygon provided by University of Yamanashi (as part of GRanD database)* *'Other' = other sources, including remote sensing imagery (e.g., MODIS) and GIS repositories (e.g., US National Hydrography Dataset)* *'No polygon' = no polygon available*	41,145 *221* *168* *314* *13,468* *986* *506* *18,887* *494* *251* *5,850*
Grand_id	Unique ID for corresponding record in GRanD database (v 1.4; Lehner et al. 2011); 0 = no record in GRanD	7,424
Hyriv_id	Unique ID for corresponding river reach in RiverATLAS database (v1.0; Linke et al. 2019); 0 = no corresponding record in RiverATLAS	41,106
Instream	Flag of whether barrier is located instream or off-stream a river reach of RiverATLAS database (v1.0; Linke et al. 2019): *'Instream' = barrier is located on a river reach of RiverATLAS* *'Offstream' = barrier is located off-stream (away from) any river reach of RiverATLAS; in that case 'Hyriv_id' identifies the reach catchment in which the barrier is located*	41,145 *31,763* *9,382*
Hylak_id	Unique ID for corresponding polygon in HydroLAKES database (v1.1; Messager et al. 2016); also corresponds to LakeATLAS database (v1.0; Lehner et al. 2022); 0 = no corresponding polygon in HydroLAKES or LakeATLAS	31,264
Hybas_L12	Unique ID for each corresponding sub-basin at level 12 in BasinATLAS database (v1.0; Linke et al. 2019); 0 = no corresponding sub-basin in BasinATLAS	41,134

Notes:

- The columns 'Area_skm' and 'Cap_mcm' have been created to provide a "most representative" estimate of reservoir surface area and reservoir storage capacity. The values were derived from other columns following the rules as indicated in Table 2. It should be noted, however, that the source values may not correctly refer to "maximum", "normal", or "minimum" conditions as this distinction was often not available in the original sources (see also next note).
- In most original data sources, no distinction was made between "maximum capacity", "gross capacity", "normal capacity", "live capacity", or "minimum capacity"; or the distinction was not reliable. If no distinction was available and only one value was reported, it was entered as 'Cap_rep'. If an explicit, reliable distinction was available, the values were entered as 'Cap_max' (for maximum or gross capacity), 'Cap_rep' (for normal capacity) and 'Cap_min' (for live or minimum capacity). If no distinction was available and two different values were reported, the most plausible one was entered as 'Cap_rep', and the other one as 'Cap_max' or 'Cap_min' according to its size. If no distinction was available and more than two values were reported, they were sorted into 'Cap_max', 'Cap_rep', and 'Cap_min' according to their size. For all records of the United States, 'Cap_max' explicitly refers to "maximum capacity" and 'Cap_rep' explicitly refers to "normal capacity".
- Regarding the use/purpose of a reservoir: 'Main' refers to the primary purpose; 'Major' refers to a primary/important purpose, yet not the main one; 'Sec' refers to a secondary purpose. Note that the distinction between reservoir purposes and their attribution as 'Main', 'Major', or 'Sec' may be uncertain and/or arbitrary in many cases as many reservoirs may have multiple/mixed purposes that are difficult to rank or determine, or that have changed over time (e.g., the main purpose of a former hydropower dam may have been superseded by recreational use today).
- Missing numerical records are flagged with value "-99"; and "-9999" for missing elevation values. Missing text records are represented as empty fields. Note that missing information does not indicate the absence of a characteristic (e.g., empty fields in the 'Dam_name', 'Main_use', or 'Use_recr' columns do not indicate that a dam has no assigned name or main purpose, nor that the reservoir is not used for recreational use; it may, in many instances, only indicate that the information is unknow in the database).

4. License, disclaimer and acknowledgement

4.1 License agreement

The Global Dam Watch (GDW) database is licensed under a Creative Commons Attribution 4.0 International License. By downloading and using the data the user agrees to the terms and conditions of this license. A copy of the license is available at http://creativecommons.org/licenses/by/4.0/. Notwithstanding this free license, we ask users to refrain from redistributing the data in whole in its original format on other websites without the explicit written permission from the authors.

4.2 Disclaimer of warranty

The Global Dam Watch (GDW) database and any related materials contained therein are provided "as is" without warranty of any kind, either express or implied, including, but not limited to, the implied warranties of merchantability, fitness for a particular purpose, noninterference, system integration, or noninfringement. The entire risk of use of the data shall be with the user. The user expressly acknowledges that the data may contain some nonconformities, defects, or errors. The authors do not warrant that the data will meet the user's needs or expectations, that the use of the data will be uninterrupted, or that all nonconformities, defects, or errors can or will be corrected. The authors are not inviting reliance on these data, and the user should always verify actual data.

4.3 Limitation of liability

In no event shall the authors be liable for costs of procurement of substitute goods or services, lost profits, lost sales or business expenditures, investments, or commitments in connection with any business, loss of any goodwill, or for any direct, indirect, special, incidental, exemplary, or consequential damages arising out of the use of the Global Dam Watch (GDW) database and any related materials, however caused, on any theory of liability, and whether or not the authors have been advised of the possibility of such damage. These limitations shall apply notwithstanding any failure of essential purpose of any exclusive remedy.

4.4 Citations and acknowledgements

The authors would like to thank the Global Dam Watch consortium and their partners for coordinating the development of the GDW database. Several international meetings and workshops were facilitated and sponsored by Worldwide Fund for Nature (WWF) Netherlands, WWF-US, and the National Socio-Environmental Synthesis Center (SESYNC) under funding received from the National Science Foundation DBI-1639145. Additional funding for the database development was provided by the World Bank, and by McGill University, Montreal, Canada. The findings, interpretations, and conclusions expressed do not necessarily reflect the views of The World Bank, its Board of Executive Directors, or the governments they represent. The authors would also to thank and acknowledge all original data providers for their invaluable contributions to this project.

Citations and acknowledgements of the Global Dam Watch (GDW) database should be made as follows:

> *Lehner, B., Beames, P., Mulligan, M., Zarfl, C., De Felice, L., van Soesbergen, A., Thieme, M., Garcia de Leaniz, C., Anand, M., Belletti, B., Brauman, K.A., Januchowski-Hartley, S.R., Lyon, K., Mandle, L., Mazany-Wright, N., Messager, M.L., Pavelsky, T., Pekel, J.-F., Wang, J., Wen, Q., Wishart, M., Xing, T., Yang, X., Higgins, J. (2024): The Global Dam Watch database of river barrier and reservoir information for large-scale applications. Scientific Data. [please insert journal volume and DOI once released]*

We kindly ask users to cite the Global Dam Watch (GDW) database in any published material produced using the data. If possible, online links to the GDW website should be provided (https://www.globaldamwatch.org).

4.5 Copyright and required attribution

The following copyright statement should be displayed with, attached to, or embodied (in a reasonable manner) in the documentation or metadata of products that are utilizing parts or all of the GDW database:

> *This product incorporates data from the GDW database © Global Dam Watch (2024).*

5. References

EEA (European Environment Agency). (2012). ECRINS (European Catchments and RIvers Network System): Lakes. Version 1. Available online at https://www.eea.europa.eu/en/datahub/datahubitem-view/a9844d0c-6dfb-4c0c-a693-7d991cc82e6e

ICOLD (International Commission on Large Dams). (1998–2022). World Register of Dams. Version updates 1998-2022. Paris: ICOLD. Available online at https://www.icold-cigb.org

Lehner, B., and Döll, P. (2004). Development and validation of a global database of lakes, reservoirs and wetlands. Journal of Hydrology 296: 1-22. https://doi.org/10.1016/j.jhydrol.2004.03.028

Lehner, B., Verdin, K., and Jarvis, A. (2008). New global hydrography derived from spaceborne elevation data. Eos 89: 93-94. https://doi.org/10.1029/2008EO100001

Lehner, B., Reidy Liermann, C., Revenga, C., Vörösmarty, C., Fekete, B., Crouzet, P., et al. (2011). High-resolution mapping of the world's reservoirs and dams for sustainable river-flow management. Frontiers in Ecology and the Environment 9: 494-502. https://doi.org/10.1890/100125

Lehner, B., Messager, M.L., Korver, M.C., and Linke, S. (2022). Global hydro-environmental lake characteristics at high spatial resolution. Scientific Data 9: 351. https://doi.org/10.1038/s41597-022-01425-z

Linke, S., Lehner, B., Ouellet Dallaire, C., Ariwi, J., Grill, G., Anand, M., et al. (2019). Global hydro-environmental sub-basin and river reach characteristics at high spatial resolution. Scientific Data 6: 283. https://doi.org/10.1038/s41597-019-0300-6

Messager, M.L., Lehner, B., Grill, G., Nedeva, I., and Schmitt, O. (2016). Estimating the volume and age of water stored in global lakes using a geo-statistical approach. Nature Communications: 13603. https://doi.org/10.1038/ncomms13603

Mulligan, M., Lehner, B., Zarfl, C., Thieme, M., Beames, P., van Soesbergen, A., et al. (2021). Global Dam Watch: curated data and tools for management and decision making. Environmental Research: Infrastructure and Sustainability 1(3): 033003. https://doi.org/10.1088/2634-4505/ac333a

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

Müller Schmied, H., Cáceres, D., Eisner, S., Flörke, M., Herbert, C., Niemann, C., et al. (2021). The global water resources and use model WaterGAP v2.2d: model description and evaluation. Geoscientific Model Development 14: 1037-1079. https://doi.org/10.5194/gmd-14-1037-2021

Natural Resources Canada. (2013). CanVec Hydrography: Waterbody Features. Version 12.0. Available online at https://open.canada.ca/data/dataset/9d96e8c9-22fe-4ad2-b5e8-94a6991b744b

Pekel, J.F., Cottam, A., Gorelick, N., and Belward, A. (2016). High-resolution mapping of global surface water and its long-term changes. Nature 540: 418-422. https://doi.org/10.1038/nature20584

Robinson, N., Regetz, J., and Guralnick, R.P. (2014). EarthEnv-DEM90: a nearly-global, void-free, multi-scale smoothed, 90m digital elevation model from fused ASTER and SRTM data. Journal of Photogrammetry and Remote Sensing 87: 57-67. https://doi.org/10.1016/j.isprsjprs.2013.11.002

Slater, J.A., Garvey, G., Johnston, C., Haase, J., Heady, B., Kroenung, G., and Little, J. (2006). The SRTM data “finishing” process and products. Photogrammetric Engineering & Remote Sensing 72: 237-247. https://doi.org/10.14358/PERS.72.3.237

USACE (US Army Corps of Engineers). 2021. National Inventory of Dams. Data available at https://nid.sec.usace.army.mil

Wang, J., Walter, B.A., Yao, F., Song, C., Ding, M., Maroof, A.S., et al. (2022). GeoDAR: georeferenced global dams and reservoirs dataset for bridging attributes and geolocations. Earth System Science Data 14: 1869-1899. https://doi.org/10.5194/essd-14-1869-2022

Yang, X., Pavelsky, T.M., Ross, M.R.V., Januchowski-Hartley, S.R., Dolan, W., Altenau, E.H., et al. (2022). Mapping flow-obstructing structures on global rivers. Water Resources Research 58: e2021WR030386. https://doi.org/10.1029/2021WR030386

Zarfl, C., Lumsdon, A.E., Berlekamp, J., Tydecks, L., and Tockner, K. (2015). A global boom in hydropower dam construction. Aquatic Sciences 77, 161-170. https://doi.org/10.1007/s00027-014-0377-0