Global Dam Watch (GDW) Database

A harmonized and curated database of river barriers and reservoirs worldwide

Technical Documentation – version 1.0 delta

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

February 2024

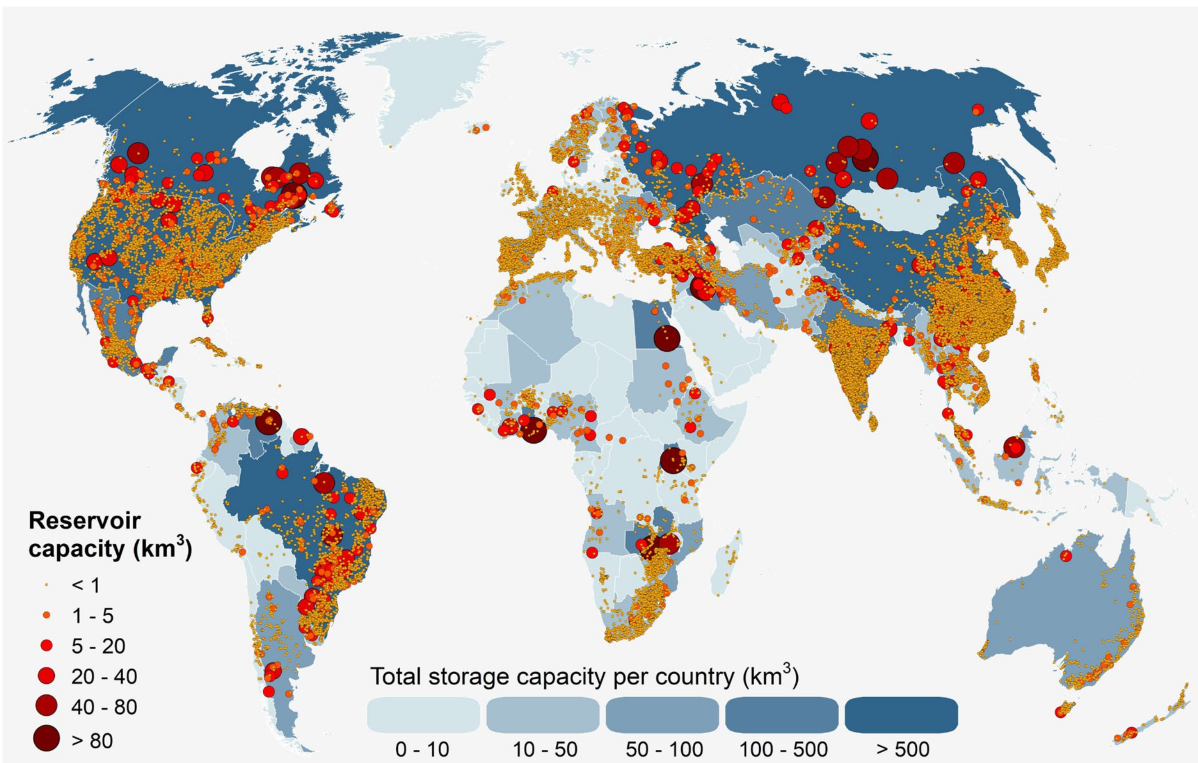

Figure 1: Global distribution of reservoirs and river barriers in GDW database.

Note:
This document refers to the pre-release delta version of the new "Global Dam Watch Database version 1", available at https://figshare.com/s/e751055c0740de16ba91
This delta version was created for the review process of the accompanying publication. Once the review process is completed, version 1.0 will be released.

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 on the planet (Lehner et al. 2011), but despite efforts by individual groups, only a small proportion of them have been mapped today.

The development of the Global Dam Watch (GDW) database is an initiative of 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 global-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 data 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 delta) contains 41,145 barrier and dam locations (Figure 1), and 35,295 associated reservoir polygons, with a cumulative storage capacity of 7405 km^3. The database is freely available for download at www.globaldamwatch.org. This Technical Documentation describes the content of the database. The development and characteristics of GDW v1 are fully described by Lehner et al. (in preparation) and can be temporarily 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., Mandle, L., Mazany-Wright, N., Messager, M.L., Pavelsky, T., Pekel, J.-F., Wang, J., Wen, Q., Xing, T., Yang, X., Wishart, M., Lyon, K., Higgins, J. (in preparation): The Global Dam Watch database of river barrier and reservoir information for large-scale applications.

2. Methods

The details of the GDW database development are described in Lehner et al. (*in preparation*). 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, ESRI, and Bing Maps. 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 at Chapel Hill, USA; Swansea University, UK; and World Wildlife Fund, Washington DC, USA.

2.1 Main data sources

The development of the harmonized GDW database primarily aimed at compiling available global barrier, dam, and reservoir information; curating it through both (supervised) automated and manual cross-validation, error checking, and identification of duplicate records, attribute conflicts, or mismatches; and completing missing information from a multitude of sources or statistical approaches. Table 1 shows the main source datasets that were used to create the first version of the GDW database. While all sources resemble global data repositories, they show different characteristics regarding their content, comprehensiveness, and the type of attributes they provide, mostly due to the different objectives and needs when assembling 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 underlying 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*	*Attributes*	*Number of contributed records**
GOODD (GlObal geOreferenced Database of Dams)	Mulligan et al. 2020	Medium to large dam locations that are visible on satellite imagery (Google Earth); dam ≥150 m or reservoir ≥500 m long; manually digitized	Point coordinates	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, literature, internet; manual inspection and validation of all records; extensive attribute information	Point coordinates, polygon outlines, multiple attributes including name, year, height, purpose, reservoir volume	7,424
FHReD (Future Hydropower Reservoirs and Dams database)	Zarfl et al. 2015	Hydropower dams ≥1MW; compiled from freely available data, literature, internet; manual inspection and validation of all records; the original dataset focused on planned projects, from which those that were completed by 2022 were selected	Point coordinates, hydropower capacity	205
JRC-GSW (Global Surface Water Explorer of European Commission's Joint Research Centre)	Pekel et al. 2016	Surface water extents mapped at 30m grid resolution from Landsat imagery; automatic extraction of new reservoirs that appear after 1983	Polygon outlines	1,451
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	Point coordinates, obstruction type	6,113
HydroLAKES	Messager et al. 2016	Polygon outlines for all 'lakes' globally with a surface area ≥10 ha; polygons were used as reservoir outlines if they were associated with a barrier/dam (from GOODD, FHReD or GROD)	-	-
HydroSHEDS	Lehner et al. 2008	River network to which the barrier/dam locations were co-registered; after co-registration, some hydrometric attributes were derived, incl. catchment area and discharge	-	-

* 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 v1.4 (Global Reservoir and Dam database; 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 prioritized major dams that form larger reservoirs, as well as instream barriers on larger rivers, for which more information is available. This focus on ‘larger’ structures already existed 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 has an exclusive focus on 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/dam (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 automatic ‘spatial join’ procedure (i.e., identified by barrier points that fell inside an existing lake polygon from HydroLAKES), or by manual inspections of candidate HydroLAKES polygons that were in close vicinity (<1 km) of barrier or dam locations. In addition, where needed, 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 some 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, each record in the GDW database—as identified by a unique ID—typically represents a combined 'reservoir-and-dam object' and is defined by both a polygon and a point location (see also section 3.1 on data formats). The point represents the location of the barrier or dam, or of the 'main' dam in case of multiple dams forming a single reservoir (these latter cases are further described in columns 'Multi_dams' and 'Comments' in the attribute table). Furthermore, barrier/dam objects can also be defined by a point only, representing an independent barrier or dam without a "traditional" reservoir, including run-of-the-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 which are available since 1984 (Pekel et al. 2016). For the creation of GDW v1, the JRC-GSW maps showing 'maximum surface water extent' were used for the period 1984-2020. 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 delineated, the polygons were manually inspected and, if necessary, corrected by consulting imagery from ESRI Basemaps, Google Maps, Yandex Maps, Mapbox, JRC-GSW change maps, NASA Worldview imagery, and other auxiliary documents pertaining to the reservoir. Where necessary, these images were georeferenced and used to manually delineate or correct reservoir shorelines. In particular, adjustments were made 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 associated barrier(s) and dam(s), which supported the identification and elimination of duplicates. If barrier or dam points from multiple source datasets were associated with the same reservoir polygon, they were considered duplicates and only one merged record was kept in the GDW database.

For barrier and dam locations without reservoirs, duplicity was less easy to detect. In iterative, 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 barrier and dam records that are represented only by points (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, ESRI, and Bing Maps). For records with a reservoir polygon, the reservoir's outlet point was used as a proxy for its dam location (see section 2.2 above), which by default is located on 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 HydroSHEDS river network, spatial offsets and uncertainties in the range of 500 m are inherent in the HydroSHEDS database due to the applied grid cell resolution of 15 arc-seconds (~500 m at the equator). Therefore, the representative barrier/dam 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 HydroATLAS 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 (Döll et al. 2003) for the period 1971-2000 and were also used to calculate the 'Degree of Regulation' index for every dam (see Table 2).

2.7 Estimating missing reservoir volumes

In the course of constructing the GDW database, two equations were derived and applied to complete missing reservoir volumes, following the approach by Lehner et al. (2011):

$$V = 0.5789\,(A \cdot h)^{0.9412} \qquad \text{(Eq. 1)}$$

$$V = 13.794\,A^{1.059} \qquad \text{(Eq. 2)}$$

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

Both equations were determined by a statistical regression analysis of 7,348 reservoirs worldwide contained in the GDW v1 database which were selected based on data reliability using the following characteristics: each record showed a reported reservoir capacity, reported dam height, and included a surface polygon; the calculated mean depth of each reservoir (reported capacity 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. Three additional records in GDW v1 matched these requirements but were dismissed as clear outliers after inspecting the regression scatter plots (each represented an extremely large but shallow reservoir). Eq. 1 was used to estimate missing reservoir volumes if both area and dam height were available (R^2

= 0.94 for reservoirs used in the determination of the equation's parameter settings); Eq. 2 was used if only the reservoir area was available (R^2 = 0.82).

Note that Eqs. 1 and 2 were derived by relating reported 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.

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, an estimate of the filling year was made using timelapse remote sensing imagery built from the Landsat archive on Google Earth Engine (see https://earthengine.google.com/timelapse/). First, a 'candidate' year was derived through a newly developed algorithm 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 of the reservoir in the timelapse sequence. Reservoirs that were already filled before the first Landsat imagery was available in 1984 were flagged as 'before 1985' (or 'before [year]' in cases where obscured imagery prevented detection before a later year).

While distinct changes in the timelapse sequences were observed for many records, some cases were ambiguous, either due to obscured imagery (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 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 pre-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 dam and reservoir was compiled and cross-referenced using multiple sources to verify veracity and identify conflicts. Where available, links to source materials were included in the respective record for reference. 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 as described above, 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 flag (i.e., *Verified, Good, Fair, Poor,* or *Unreliable*) was assigned to each record (for more details see Table 2). Although subjective, this indicator allows identification of records where obvious inconsistencies, uncertainties, or data gaps remain.

Despite these 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 spatial location of the barrier, dam, or reservoir, or to its associated attribute information. For example, many inconsistencies in the GRanD database were caused by typos and order-of-magnitude errors, such as mistyped volumes by a factor of 1000; or by unit mismatches (e.g., feet vs. meters)

(Lehner et al. 2011). Also, in many cases 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 dams and reservoirs: some dams, such as barrages, diversions, or run-of-river hydropower stations, may not form reservoirs; some impoundments may have multiple dams (e.g., main and saddle dams); and some reservoirs have no dams 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 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 identified by GeoDAR and corrected accordingly 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_delta' is a point layer containing all estimated dam and barrier locations and their attribute information
- 'GDW_reservoirs_v1_delta' is a polygon layer containing all corresponding reservoir outlines and their attribute information

Each dam 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'). Version 0.2 of the GDW database contains 41,145 barrier points and 35,295 reservoir polygons. I.e., 5,850 barrier locations have no associated reservoir 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 are provided in ESRI© Geodatabase and Shapefile formats. Each shapefile consists of five core files (.dbf, .sbn, .sbx, .shp, .shx); and projection information is provided in an ASCII text file (.prj). The data are provided in an unprojected Geographic Coordinate System using the World Geodetic System 1984 (GCS_WGS_1984). The GDW database is available electronically in compressed zip file format. To use the data files, the zip files must first be decompressed. Each zip file includes a copy of the GDW Technical Documentation.

NOTE: Users without GIS software or without the option to interpret shapefiles may import the .dbf file (in dBASE IV format) in most spreadsheet programs. This file contains all GDW attribute information, and the dam 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).*

Column title	*Description*	*Number of occurrences*
GDW_ID	Unique ID for each barrier/dam and associated reservoir; IDs correspond between barrier (point) and reservoir (polygon) layers of the GDW database	41,145
Grand_id	Unique ID for each original record in the GRanD database (version 1.4)	7,424
Res_name	Name of reservoir or lake (i.e., impounded water body)	2,098
Dam_name	Name of dam structure	10,071
Alt_name	Alternative name of reservoir or dam (including different spelling, different language)	1,807
River	Name of impounded river	9,501
Alt_river	Alternative name of impounded river (including different spelling, different language)	714
Main_basin	Name of main basin	2,738
Sub_basin	Name of sub-basin	721
Near_city	Name of nearest city	6,370
Alt_city	Alternative name of nearest city (including different spelling, different language)	302
Admin_unit	Name of administrative unit	8,139
Sec_admin	Secondary administrative unit (indicating dams or reservoirs that lie within or are associated with multiple administrative units)	85
Country	Name of country	41,145
Sec_cntry	Secondary country (indicating international dams or reservoirs that lie within or are associated with multiple countries)	76
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; 'Enlarged' = lake control structure enlarged the original lake surface area; 'Maybe' = not sure, but data seems to indicate a lake control structure	209
Year_dam	Year in which the dam or 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 column 'Year_src')	15,230
Pre_year	Estimated year before which the barrier was built (e.g., 1985 in this column means 'pre-1985' or '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	2,518
Year_src	Source of information for 'Year' or 'Pre-year': reported in 'GRanD', 'NID' (USACE 2021), or 'Other'; derived through AI-supported auto-detection using 'JRC-GSW' data (Pekel et al. 2016); or 'Estimated' by analyzing time-lapse data of remote sensing imagery	17,749
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 was removed, replaced, subsumed, or destroyed; see also column 'Timeline' below	10

Column title	*Description*	*Number of occurrences*
Timeline	Indicates whether the status of a dam has changed or will 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 and point are retained but the dam itself has been removed and not replaced) 'Replaced' (dam record and point are retained in dataset but the dam itself has been replaced; the new dam is recorded as a new point) 'Subsumed' (dam record and point are retained but the dam and reservoir themselves were subsumed by larger infrastructure constructed further downstream; the old reservoir polygon has been removed in version 1.3 and the new dam and reservoir are recorded as a new point and polygon) 'Under construction' (dam is currently under construction)	70
Year_txt	Summary of year information in text format	41,145
Dam_hgt_m	Height of dam in meters	9,311
Alt_hgt_m	Alternative height of dam (may indicate update or secondary dam construction)	366
Dam_len_m	Length of dam in meters	8,276
Alt_len_m	Alternative length of dam (may indicate 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'; exceptions apply if value in 'Area_poly' column seems unreliable; see also notes below	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'; 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 notes below	35,334
Cap_max	Reported 'maximum storage capacity' in million cubic meters; see notes below	4,403
Cap_rep	Reported 'storage capacity' in million cubic meters; value may refer to different types of storage capacity; see notes below	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
Dis_avg_ls	Long-term (1971-2000) average discharge at dam location in liters per second; value derived from HydroSHEDS flow routing scheme combined with WaterGAP runoff estimates (Döll et al. 2003) at 15s resolution at point location of barrier/dam	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 data set (Robinson et al. 2014) at 15s resolution as minimum within reservoir polygon or at point location of barrier/dam, respectively	41,134

Column title	*Description*	*Number of occurrences*
Catch_skm	Area of upstream catchment draining into the reservoir in square kilometers; value derived from HydroSHEDS at 15s resolution at point location of barrier/dam	41,134
Catch_rep	Reported area of upstream catchment draining into reservoir in square kilometers	4,007
Power_mw	Hydropower capacity in MW	240
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
Use_irri	Used for irrigation ('Main'; 'Major'; 'Sec' = Secondary use; or 'Multi' if multiple uses exist without a ranking)	2,669
Use_elec	Used for hydroelectricity production ('Main'; 'Major'; 'Sec'; or 'Multi')	3,064
Use_supp	Used for water supply ('Main'; 'Major'; 'Sec'; or 'Multi')	2,285
Use_fcon	Used for flood control ('Main'; 'Major'; 'Sec'; or 'Multi')	2,030
Use_recr	Used for recreation ('Main'; 'Major'; 'Sec'; or 'Multi')	2,104
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)	8,435
Multi_dams	Indicates whether there is more than one dam 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	1,227
Quality	Quality index: 1: Verified (location and data have been verified) 2: Good (location and data seem good but have not all 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
Editor	Final data editor: 'McGill' = McGill University (BL = B. Lehner; PB = P. Beames; MA = Mira Anand; TX = Tianqi Xing) 'UNH' = University of New Hampshire	41,145
Long_riv	Longitude of the point location of the barrier/dam 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 barrier/dam 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 barrier/dam in decimal degrees after it was associated with a river segment of HydroSHEDS; see associated 'Long' column for more details	41,145
Long_dam	Longitude of the actual point location of the barrier/dam in decimal degrees; i.e., this represents the actual location of the barrier/dam 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 barrier/dam location	6,113
Lat_dam	Latitude of the actual point location of the barrier/dam in decimal degrees; see associated 'Long' column for more details	6,113

Column title	*Description*	*Number of occurrences*
Orig_src	Original dataset from which the barrier/dam or reservoir was derived: 'GRanD' = Global Reservoir and Dam database v1.4 (Lehner et al. 2011) 'GOODD' = GlObal geOreferenced Database of Dams (Mulligan et al. 2020) 'GOODD-NID' = GOODD with attribute information from NID (USACE 2021) 'FHReD' = Future Hydropower Reservoirs and Dams Database (Zarfl et al. 2015) '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
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 and Döll 2004) 'JRC-GSW' = polygon digitized from European Commission Joint Research Centre's Global Surface Water Explorer data (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 'Other' = other sources, including remote sensing imagery (e.g., MODIS) and GIS repositories (e.g., US National Hydrography Dataset, US Geological Survey 2013) 'No polygon' = no polygon available	41,145
Hylak_id	Unique ID for each corresponding polygon in the HydroLAKES database (version 1.1; Messager et al. 2016); also corresponds to the LakeATLAS database (version 1.0; Lehner et al. 2022)	31,264
Hyriv_id	Unique ID for each corresponding river reach in the RiverATLAS database (version 1.0; Linke et al. 2019)	41,106
Instream	'Instream' = barrier is located on a river reach of RiverATLAS 'Offstream' = barrier is located offstream (away from) any river reach of RiverATLAS; in that case 'Hyriv_id' identifies the reach catchment in which the barrier is located	41,145
Hybas_l12	Unique ID for each corresponding subbasin at level 12 in the BasinATLAS database (version 1.0; Linke et al. 2019)	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 (note that the distinction between 'Main' and 'Major' may be arbitrary in some cases); 'Sec' refers to a secondary purpose.
- Missing numerical records are flagged with value "-99"; and "-9999" for missing elevation values.

4. License, disclaimer and acknowledgement

4.1 License agreement

Version 1 of 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. GDW v1 is available for public download at https://www.globaldamwatch.org.

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 WWF Netherlands, the WWF Innovation Fund, 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 like to thank and acknowledge all original data providers for this project.

Citations and acknowledgements of the Global Dam Watch (GDW) database should be made as follows (note that the following citation is only temporary until final publication):

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., Mandle, L., Mazany-Wright, N., Messager, M.L., Pavelsky, T., Pekel, J.-F., Wang, J., Wen, Q., Xing, T., Yang, X., Wishart, M., Lyon, K., Higgins, J. (in preparation): The Global Dam Watch database of river barrier and reservoir information for large-scale applications.

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 GLWD 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:

5. References

Döll, P., Kaspar, F., and Lehner, B. 2003. A global hydrological model for deriving water availability indicators: model tuning and validation. Journal of Hydrology 270: 105–134.

EEA (European Environment Agency). 2012. ECRINS (European Catchments and RIvers Network System): Lakes. Version 1. Available online at http://projects.eionet.europa.eu/ecrins.

ICOLD (International Commission on Large Dams). 1998–2018. World Register of Dams. Version updates 1998-2009. Paris: ICOLD. Available online at www.icold-cigb.net.

Lehner, B., and Döll, P. 2004. Development and validation of a global database of lakes, reservoirs and wetlands. Journal of Hydrology 296: 1–22.

Lehner, B., Verdin, K., and Jarvis, A. 2008. New global hydrography derived from spaceborne elevation data. Eos 89: 93–94.

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

Lehner, B., Messager, M.L., Korver, M.C., 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., Beames, P., Burchard-Levine, V., Maxwell, S., Moidu, H., Tan, F., Thieme, M. 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., 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

Natural Resources Canada. 2013. CanVec Hydrography: Waterbody Features. Version 12.0. Available online at ftp://ftp2.cits.rncan.gc.ca/pub/canvec/.

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.

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.

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

US Geological Survey. 2013. National Hydrography Geodatabase: Alaska. Available online via The National Map viewer at http://viewer.nationalmap.gov/viewer/nhd.html?p=nhd or at http://nhd.usgs.gov/.

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., Tockner, K. (2015). A global boom in hydropower dam construction. Aquatic Sciences 77, 161-170. https://doi.org/10.1007/s00027-014-0377-0