SCIENTIFIC DATA

Check for updates

OPEN

DATA DESCRIPTOR

A socio-environmental geodatabase for integrative research in the transboundary Rio Grande/Río Bravo basin

Sophie Plassin [1], Jennifer Koch [1 ✉], Stephanie Paladino[2,3], Jack R. Friedman[3], Kyndra Spencer[1] & Kellie B. Vaché[4]

Integrative research on water resources requires a wide range of socio-environmental datasets to better understand human-water interactions and inform decision-making. However, in transboundary watersheds, integrating cross-disciplinary and multinational datasets is a daunting task due to the disparity of data sources and the inconsistencies in data format, content, resolution, and language. This paper introduces a socio-environmental geodatabase that transcends political and disciplinary boundaries in the Rio Grande/Río Bravo basin (RGB). The geodatabase aggregates 145 GIS data layers on five main themes: (i) Water & Land Governance, (ii) Hydrology, (iii) Water Use & Hydraulic Infrastructures, (iv) Socio-Economics, and (v) Biophysical Environment. Datasets were primarily collected from public open-access data sources, processed with ArcGIS, and documented through the FGCD metadata standard. By synthesizing a broad array of datasets and mapping public and private water governance, we expect to advance interdisciplinary research in the RGB, provide a replicable approach to dataset compilation for transboundary watersheds, and ultimately foster transboundary collaboration for sustainable resource management.

Background & Summary

Finding solutions to manage scarce water resources to meet human water needs while sustaining ecosystems has become a priority for research[1–4] and policy makers. In 2015, the United Nations (UN) adopted 17 Sustainable Development Goals (SDG), including "Ensuring access to water and sanitation for all" (SDG6)[5]. However, increasing pressure on water resources, due to population growth and climate change, may be an obstacle to reaching this goal and may lead to increased competition and political tensions. This especially applies to transboundary river systems[6], which represent 60% of freshwater supply basins and almost 50% of the total land area on Earth[7].

In transboundary basins, data- and information-sharing provide a mechanism to foster cooperation among countries and, ultimately, achieve more sustainable water management and maintain peace and security[7]. Synthesizing a broad range of environmental and social datasets is further critical for in-depth understanding of watershed dynamics[8] and for assessing the sustainability of water allocation[9]. In large areas, spatial datasets are also valuable to capture the spatial heterogeneity of local conditions and for better understanding of different development trends across a region[10]. However, integrating scattered cross-disciplinary and multinational datasets is challenged by the disparity of data sources, transboundary discontinuity[11] and the inconsistencies in format, content, spatial and temporal resolution, languages[12], and institutional norms. This can limit the study of transboundary socio-environmental systems, and calls for the development of a harmonized, transboundary, open-access database.

[1]University of Oklahoma, Department of Geography and Environmental Sustainability, 100 E Boyd St, Suite 510, Norman, OK, 73019, USA. [2]MeroLek Research, PO Box 973, Athens, 30603, GA, USA. [3]University of Oklahoma, Center for Applied Social Research, 5 Partners Place, 201 Stephenson Parkway, Suite 4100, Norman, 73019, OK, USA. [4]Oregon State University, Biological and Ecological Engineering, 124 SW 26th Street, Gilmore Hall 232, Corvallis, 97331, OR, USA. ✉e-mail: jakoch@ou.edu

The Rio Grande/Río Bravo basin (RGB) epitomizes the challenges of managing water resources under conditions of scarcity[13] and transboundary cooperation issues[14]. Covering 552,382 km², the basin is shared by two countries (United States and Mexico) and eight states, three in the U.S. (Colorado, New Mexico, and Texas) and five in Mexico (Chihuahua, Coahuila, Durango, Nuevo León, and Tamaulipas). Distribution of water resources is governed by distinct national and state water rights regimes, as well as bi-national and inter-state agreements[14]. The river is highly engineered via extensive damming and channelization[15], which significantly disturb the natural flow regime[16]. Classified as one of the most endangered in the world[17], the river hosts several endangered and threatened species, such as the Rio Grande Silvery Minnow, and the Southwestern Willow Flycatcher. Despite chronically severe water scarcity, human water demand has grown along with an increase in population (current population is estimated at 10.5 million people[18]). Furthermore, irrigated agriculture makes up approximately 83.5% of the total surface and groundwater withdrawals[19,20] while cultivated areas cover only 3.5% of the basin[21]. Projected climate change and population growth are likely to lead to a growing imbalance between water supply and demand, potentially impacting the sustainability of the basin[22] and resulting in a cascade of negative impacts (e.g., increased risk of wildfires)[23].

To overcome the challenges of data- and information sharing in transboundary basins, several initiatives were launched in recent years. The Transboundary Freshwater Dispute Database gathers global and regional information for the world's 310 international river basins, including the RGB[24]. Examples of topics include population, land-cover, irrigation, rainfall, and dams (http://gis.nacse.org/tfdd/index.php). In the RGB, a basin level database synthesizing hydrologic, administrative, land-use/cover, and water management datasets, was also developed at a finer resolution[25,26]. However, we are not aware of a similar approach to capture the social heterogeneity and complexity across the basin[27]. Hence, we compiled a comprehensive socio-environmental geodatabase for the RGB, encompassing geospatial data sets related to *Water & Land Governance*, *Hydrology, Water Use & Hydraulic Infrastructures*, *Socio-Economics*, and *Biophysical Environment*. An innovative contribution of our geodatabase is to provide open-access to a thorough collection of reliable and well documented, basin-wide geospatial data sets, as well as, to our knowledge, the most extensive spatial picture of multi-scale water governance in the RGB. The geodatabase expands the availability of information detailing the social and environmental characteristics of the basin and enables an integrated and cross-disciplinary approach to the watershed. Many outputs can be derived from the geodatabase, including maps, spatially explicit integrated models, and statistical analyses (see *Usage Notes* for additional details). The data synthesis approach presented here is transferable to other transboundary basins, and the geodatabase will help the research, management, and policy-making communities to foster transboundary collaboration.

Methods

Dataset preparation.   The development of the socio-environmental geodatabase followed four steps: (1) data identification; (2) collection of raw data; (3) assembling, harmonizing, and geoprocessing of datasets; and (4) organization of the final geospatial data sets into the geodatabase (Fig. 1).

First, we drew on an interdisciplinary collaboration among social and environmental scientists to identify critical data requirements to support socio-environmental research in the RGB. The environmental anthropologists shared primary data analysis from ethnographic research and developed a qualitative typology of actors that we used to identify water and land governance datasets. We integrated land management as a key component for the study of socio-environmental dynamics in the watershed because land management decisions affect water run-off, stream flow, aquifer recharge, and soil erosion. The actor typology distinguishes the actors according to their roles in water and land management (water supply, consumption, infrastructure management, environmental protection, recreation) and their spatial scale of action (from local to bi-national level; Supplementary Information 1).

Second, we gathered 210 raw datasets from 49 distinct sources totalling around 53 GB. Where possible, and if the spatial resolution was high enough, we prioritized global datasets over national datasets and national datasets over regional datasets to limit data inconsistency and heterogeneity. Most of the raw datasets were collected from national agencies (67%), followed by state agencies (13%), global agencies (2%) and binational agencies (7%). We also collected datasets from joint research projects (5%) and private institutions (6%). Most of the input datasets were available open-access online. We also contacted public or private institutions to access unpublished information and were granted permission to publish the synthesized data. The original formatting of the raw datasets varied greatly. 62% were vector data (51.5% shapefiles, 3% KML/KMZ files, 7.5% feature class), 19% were raster files, and 19% were only available in tabulated format (comma-delimited text or Microsoft Excel files). In terms of language, 73% of the raw datasets were in English and 27% in Spanish. The detailed list of raw datasets used to produce the geospatial output datasets is available in Supplementary Information 2. For each raw dataset, we included a short description, the spatial and temporal dimensions, the format, language, data source, and date of access.

Third, we assembled, harmonized, and geoprocessed the datasets using ESRI's ArcGIS 10.6 and, where appropriate, the Python programming language (version 2.7) and the arcpy package for automating the geoprocessing workflow. Geoprocessing operations varied according to the type of raw dataset, and a detailed description of these operations is available in the metadata for each geographic layer. In general, we tried to synthesize scattered state and national raw data into one data layer. Where necessary, we harmonized the information in the datasets (variables, values, units of measurement) and translated Spanish information to English. Geospatial operations included the projection of the final datasets to the most appropriate projected coordinate system that minimized areal distortion for the whole RGB (North America Albers Equal Area Conic), and clipping datasets to the RGB boundary. Instead of using the hydrological basin boundary, we used the spatial boundaries of the Rio Grande/Río Bravo socio-environmental system as delineated by Koch *et al.*[27]. For all political jurisdictional boundary information (national, state, or county), given the spatial mismatch with the catchment boundary, we prioritized the use of administrative boundaries over catchment boundaries in locations where they overlap.
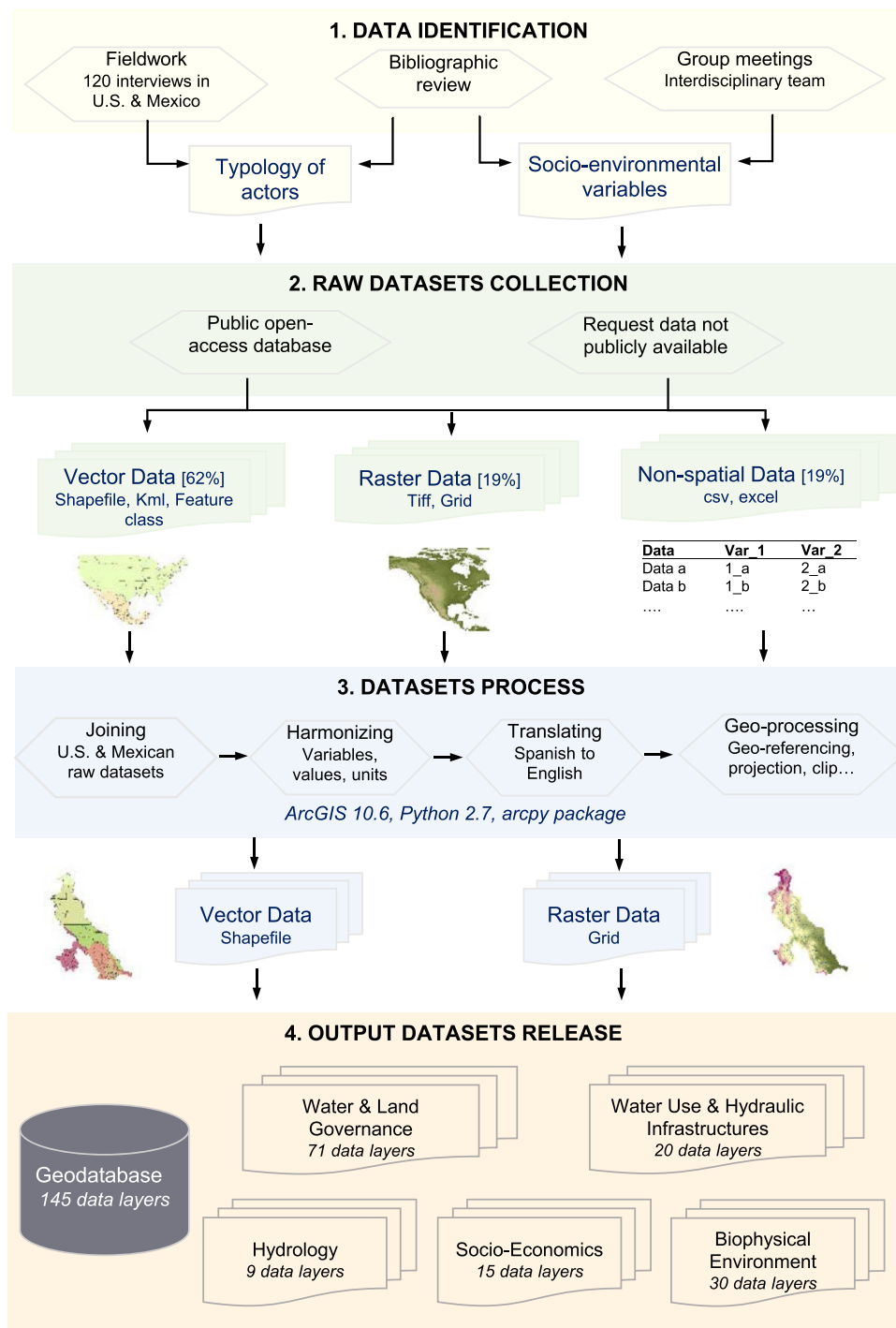
Fig. 1 Schematic overview of the workflow applied to produce the socio-environmental geodatabase for the Rio Grande/Río Bravo basin.

Fourth, we stored the resulting 145 geographic layers, including 125 vector data and 20 raster data (GRID), in a geodatabase with a total size of 1.40 GB. We organized the datasets into five main themes: *Water & Land Governance*, *Hydrology, Water Use & Hydraulic Infrastructures*, *Socio-Economics*, and *Biophysical Environment*.

Metadata. For each individual data layer, we produced a standardized metadata record, following the Federal Geographic Data Committee Content Standard for Digital Geospatial Metadata (FGDC CSDGM). The standard includes seven types of information: Identification Information (basic information about the dataset such as the citation, description, spatial domain, access constraints, etc.), Data Quality Information, Spatial Data Organization Information, Spatial Reference Information (Geographic Coordinate System Definition), Entity and Attribute Information, Distribution Information, and the Metadata Reference Information. The metadata

Datasets	Output geospatial layers	Description	Format	Input source
Watershed boundary	Sub_Basins RGB_Basin	Watershed boundaries	polygon	FAO[34]
River network	Rivers	River network	line	FAO[35]
Aquifer boundaries	Aquifer	Aquifer boundaries in U.S. and Mexico	polygon	USGS[36], CONAGUA[37]
Gauging stations	1. Gauges_usgs 2. Gauges_ibwc 3. Gauges_conagua	1. Gauges operated by USGS in U.S. 2. Gauges operated by IBWC along the border 3. Gauges operated by CONAGUA in MX	points	USGS[38], USIBWC[39], CONAGUA[40]
Quality Stations	Quality_Stations	Quality stations managed by IBWC and TCEQ	point	USIBWC[39]

Table 1. List of geospatial layers for the *Hydrology* category including their description, format and input data sources.

records have been created through the ArcMap graphical user interface and are automatically included with the hosted layer item. The user can export the metadata in XML format through the ArcMap graphical user interface if needed.

Data Records

The socio-environmental geodatabase is freely and publicly available on the OSF repository[28]. The collection of datasets can be downloaded as a single file (.gdb) or as individual geospatial output layers (shapefiles or raster files). Here, we describe the geospatial datasets available for the five themes *Water & Land Governance*, *Hydrology, Water Use & Hydraulic Infrastructures*, *Socio-Economics*, and *Biophysical Environment*. Online-only Table 1 and Tables 1–4 provide a systematic overview of the individual layers, i.e., the name, description, format and data sources for each theme. A comprehensive list of acronyms is included in Supplementary Information 3.

Water & Land Governance (GOV). The Water & Land Governance category is divided into 12 sub-categories, representing political jurisdiction boundaries, surface water management agencies (binational, federal, state), state and interstate multi-stakeholder platforms, groundwater-focused institutions, irrigation and conservation organizations, land management, and border control. A description of each entity with their missions is provided in Supplementary Information 1.

Political jurisdiction boundaries. The geodatabase provides the boundaries of four political jurisdiction divisions: nation, state, county/*municipio*, and places (cities, towns and villages). The geodatabase also provides a point dataset of the most populated places in both countries, which includes the main U.S.-MX sister cities along the border.

Binational surface water management agencies. The geodatabase gathers 13 vector datasets of binational infrastructure that spatializes the joint operations of the U.S. International Boundary and Water Commission (IBWC) and its Mexican counterpart the Comisión Internacional de Límites y Aguas (CILA), regarding national allocation of water, dam and reservoir operations, river course management (the banks, the berms, the vegetation), water quality, sanitation, and flood control in the border region.

Federal surface water management agencies. This sub-category includes two datasets of federal dams. One gathers all federal dams operated by the Comisión Nacional del Agua (CONAGUA) in Mexico. The other maps all dams owned by a federal agency in the U.S., including the U.S. Army Corps of Engineers (USACE), the Bureau of Indian Affairs (BIA), the Bureau of Land Management (BLM), the Bureau of Reclamation (USBR), the U.S. Fish and Wildlife Service (USFWS) and the U.S. Forest Service (USFS).

State and intrastate surface water management agencies. This sub-category depicts the boundaries of the Offices of the State Engineer (OSE) and their regional offices (divisions and districts) in charge of the administration of water rights and interstate compacts in the U.S. For Mexico, we generated spatial layers representing state-level offices of the federal agency CONAGUA. We also included entities that engage in planning, coordination, and oversight of domestic water supply and wastewater treatment state-wide: in Chihuahua, the *Junta Central de Agua y Saneamiento* (JCAS); in Coahuila, the *Comisión Estatal de Aguas y Saneamiento* (CEAS); in Durango, the *Comisión de Agua del Estado de Durango* (CAED); in Nuevo León, the *Servicios de Agua y Drenaje de Monterrey* (SADM); and in Tamaulipas, the *Comisión Estatal del Agua en Tamaulipas* (CEAT).

State and inter-state multi-stakeholder platforms. The geodatabase includes five platforms:

- The Rio Grande Compact Commission — U.S. inter-state platform — monitors water flows and administers water-sharing among the states of Colorado, New Mexico, and Texas. We mapped its domain through three spatial layers: one representing the spatial scope of the Interstate Compacts (including the Rio Grande Compact), one identifying the gauges used to monitor and account for states' water delivery obligations to each other, and one identifying the "post-Compact" reservoirs used to manage states' water delivery obligations.

Datasets	Output geospatial layers	Description	Format	Input source
Withdrawals	Withdrawals_2015	Amount of surface water and groundwater used in mcm by the different sectors on a county/*municipio* basis in 2015	polygon	USGS[41], CONAGUA[37]
Irrigation	1. Irrigation_US 2. Irrigation_MX	Spatial coverage of irrigated lands and irrigation systems (surface, sprinkler micro-irrigation, other) on a county/*municipio* basis	polygon	USGS[41], INEGI[42]
Water Rights	1. CO_WaterRights 2. NM_WaterRights 3. TX_WaterRights 4. COA_WaterRights 5. CHI_WaterRights 6. DUR_WaterRights 7. NVL_WaterRights 8. TAM_WaterRights	1. Surface and Ground Water rights in Colorado (Division 3) 2. Surface and Ground Water rights in New Mexico 3. Surface Water rights in Texas 4. Surface water & Ground Water rights in Coahuila 5. Surface water & Ground Water rights in Chihuahua 6. Surface water & Ground Water rights in Durango 7. Surface water & Ground Water rights in Nuevo León 8. Surface water & Ground Water rights in Tamaulipas	point	CDWR[43], NMOSE[44], TCEQ[45], CONAGUA[46]
Dams	1. Dams_us 2. Dams_mx	1. All dams in U.S. derived from USACE-NID 2. All dams in MX derived from CENAPRED	point	USACE[47], CENAPRED[48]
Water diversion structures	1. Diversion_US 2. Diversion_MX 3. Diversion_CO	1. Water diversion (canal/ditch, connector, pipeline, underground conduit) in U.S. 2. Water diversion (canals, aqueducts) in MX 3. Canals in CO	line	USGS-NHD[49], INEGI[50], CWCB/DWR[51]
Wells	1. Wells_us_nhd 2. Wells_co_cwcb 3. Wells_nm_nmose 4. Wells_tx_twdb	1. Wells in U.S. derived from USNHD 2. Wells in CO compiled by CWCB/DWR 3. Wells in NM compiled by NMOSE 4. Wells in TX compiled by TWDB	point	USGS-NHD[49], CWCB/DWR[51], NMOSE[44], TWDB[52]

Table 2. List of geospatial layers for the *Water Use & Hydraulic Infrastructures* category including their description, format and input data sources.

- The San Juan Chama Project (SJCP) — U.S. inter-state platform — moves water through a trans-mountain diversion from the Colorado River basin in the state of Colorado (San Juan River) to municipal and irrigation users in the upper and middle Rio Grande valley of New Mexico. We mapped this project with three vector datasets: one for the trans-mountain diversion, one for the reservoirs where imported water is stored and transferred, and one displaying the users of that water (counties, municipalities, irrigation districts, tribe).
- The Rio Grande Project — binational and inter-state platform — supplies water for irrigation from the Elephant Butte Reservoir to several U.S. and Mexican irrigation districts. Its domain is represented through two vector datasets: one representing the dams used for storage (Elephant Butte, Caballo) plus the diversion dams for distributing water in the U.S. and Mexico (Percha, Leasburg, Mesilla, American, Riverside — which is inactive, International); and one representing the irrigation districts supplied by the Rio Grande Project (EBID, EPWID#1 and DR 009 Valle de Juarez).
- The *Consejo de Cuenca del Río Bravo* (Rio Bravo Watershed Council) — Mexican inter-state platform — is a multi-state entity that develops and implements a basin management plan for the CONAGUA Hydrological-Administrative region VI in Mexico. We mapped its domain using its administrative boundary.
- The Rio Grande Basin Roundtable in southern Colorado — intra-state platform — regroups various stakeholders such as counties, municipalities, and conservancy districts to identify consensual strategies that meet competing needs in the region, and to provide support and funding for basin projects[29]. The geodatabase maps the boundary of its domain of action.

Groundwater-focused institutions. This dataset maps the local entities and related infrastructures associated with the administration, management, monitoring and conservation of groundwater in the U.S. states of Colorado, New Mexico, and Texas.
- In Colorado, we mapped three levels of groundwater management: the Rio Grande Water Conservation District (RGWCD) that manages groundwater at the regional level (San Luis Valley); the Groundwater Management Subdistricts that manage groundwater at the local level (subdistrict); and the high capacity wells with augmentation/replacement plans. For the regional level, we provided the spatial domain of the RGWCD, the location of unconfined and confined wells that the RGWCD monitors, and the boundaries of the Closed Basin, an unconfined groundwater salvage project maintained by RGWCD that transfers water from the aquifer to the surface water basin.

- The state of New Mexico is divided into 39 declared groundwater basins where the OSE assumes jurisdiction over the appropriation and use of groundwater. We mapped the 18 basins overlapping the RGB.
- In Texas, we mapped the domains of the Groundwater Management Areas (GMA), the Groundwater Conservation Districts (GCD) in charge of the implementation of groundwater management plans within the GMA, and the Priority Groundwater Management Areas (PGMA) located in areas with critical groundwater problems.

In Mexico, except for the groundwater permits (Títulos y permisos de aguas nacionales) available in the section Water Use & Hydraulic Infrastructures > Water Rights/Permits, we did not find any comparable administrative or governance mechanism explicitly set up on a spatial basis to manage groundwater per se.

Category	Output geospatial layers	Description	Format	Input source
Population	Population	Population in 1990, 2000, 2010 [county/*municipio*-based]	polygon	US Census Bureau[53], INEGI[54]
Population density	popdens_00 popdens_05 popdens_10 popdens_15 popdens_20	Population density in 2000, 2005, 2010, 2015, 2020 [30 arc-second resolution]	raster	CIESIN - Columbia University[55]
Income	1. Income_US 2. DistributionIncome_MX90 3. DistributionIncome_MX00 4. DistributionIncome_MX10	1. Personal income in U.S. from 1969 to 2015 [county-based] 2.3.4. Income Distribution in MX [*municipio*-based] in 1990, 2000, 2010	polygon	USBEA[56], INEGI compiled by CONABIO[57]
Number and size of farms	1. Farms_US2007 2. Farms_US2012 3. Farms_MX2007	1.2. Number of farms, farm area (in hectares and acres) and farm area distributed by farm size in US, in 2007–2012 [county-based] 3. Number of farms and farm areas (hectares) in MX in 2007 [*municipio*-based]	polygon	USDA-NASS[58], INEGI[42]
Transport Infrastructures	1. Roads 2. Railroads	1. Roads 2. Railroads	line	CIESIN - Columbia University/ ITOS - University of Georgia[59], INEGI/NRCan/USGS/CEC[60]

Table 3. List of provided geospatial layers for the *Socio-Economics* category including their description, format and input data sources.

At the transboundary level, the U.S. and Mexico also established a binational agreement, the Transboundary Aquifer Assessment Program (TAAP), to strengthen collaborations among Mexican and U.S. institutions to jointly assess priority shared aquifers along the U.S.-Mexico border[30]. Two of the four transboundary priority aquifers underlay the RGB: Mesilla/Conejos-Médanos and Hueco Bolson.

Irrigation distribution organizations. This sub-category maps the boundaries of the organizations that divert and coordinate surface water delivery to irrigators, and ensure maintenance of the irrigation conveyance systems, in both the U.S. and Mexico. Our dataset provides the name and the area managed by the organizations, in addition to other specific information for each state and country. We sought to gather the spatial boundaries of three organizations: the irrigation districts (*Distritos de Riego* in Mexico), ditch companies, and community ditch associations, an umbrella term under which we include the *acequia* systems of Colorado and New Mexico, and *Unidades de Riego* (*Irrigation Units*) across the states of Mexico. However, spatial datasets of ditch companies and community ditch associations were unavailable for New Mexico and Mexico.

Land management. This dataset maps four types of land management: Certified Ejido & Communal Land, Public, Native American/Tribal, and Private/Not Reported. The dataset results from the compilation of two national datasets: the National Surface Management Agency Area Polygons released by the Bureau of Land Management (BLM) that maps state and federally owned lands as well as Native American tribal lands in the U.S., and the *Perimetrales de los núcleos agrarios certificados* produced by the *Registro Agrario Nacional* (RAN) that maps *ejido* and *comunidad* (communal) lands in Mexico. When compiling the datasets, we labelled all areas that were neither classified by the BLM nor the RAN as "Private/Not reported".

Protected areas. This dataset maps the areas designated as protected by the International Union for Conservation of Nature (IUCN).

Border control. This sub-category maps three key components of border control along the international reach of the Rio Grande/Río Bravo: the U.S. Southwest Border Patrol Sectors associated with the migration apprehension statistics; the border crossing ports of entry; and vehicle and pedestrian fences/barriers. U.S. border control is very active and has important implications for the management of riverbank and riparian areas. However, this research to-date has not covered policies or practices by Mexican border-related agencies that affect water management or flow.

Soil and water conservation districts (included only for the U.S.). The vector dataset provided in this sub-category map the outlines of the Soil and Water Conservation Districts (SCWDs) in Colorado, New Mexico, and Texas. SCWDs foster voluntary conservation practices among private and public landowners, helping to manage and protect soil, water, forests, and wildlife at the local level. The SWCDs are a specific kind of district, with a mandate that is particular to the legal and institutional history of land and water management in the U.S. Our research to-date has not found anything equivalent to this formation in Mexico, and even less so, any spatial datasets for it.

Collaborative conservation projects. This sub-category includes conservation projects where members of multiple sectors and/or institutions actively work in collaboration to identify, protect, manage, monitor, and/or develop policy for the conservation of specific sites. Our research to-date has not extended to documentation of all such projects in either the U.S. or Mexico, and did not uncover spatial data sets that document their location. Our datasets map three collaborative conservation projects. Two of them, the Landscape Conservation Cooperatives (LCCs) and the North America Bird Conservation Joint Ventures (BJVs), are transboundary. For both, we mapped the boundaries of the four LCCs and the five BJVs that span the RGB. While the U.S. Fish and

Datasets	Output geospatial layers	Description	Format	Input source
Ecoregions	1. NA_ECO_3 2. US_ECO_4	1. Ecoregions Level III for U.S. and Mexico 2. Ecoregions Level IV for U.S.	polygon	US-EPA[61,62]
Habitat	1. US_Critbab_Line, US_Crithab_Poly 2. US_Riparian 3. US_Wetlands 4. MX_Ramsar_Site 5. MX_SAP 6. MX_SPR	1. U.S. FWS Threatened & Endangered Species Active Critical Habitat Boundaries 2. Riparian areas in the U.S. 3. Wetlands in the U.S. 4. Ramsar sites in Mexico 5. Priority Focus Sites for Biodiversity Conservation in Mexico 6. Priority Sites for Restoration in Mexico	polygon, line	USFWS[63–65], CONANP[66], CONABIO[67,68]
Soil Cover	soil	Soil units based on FAO taxonomy	polygon	FAO[69,70]
Elevation	elev	Elevation derived from the SRTM DEM and the USGS GTOPO30 [30 arc-second resolution]	raster	FAO/IIASA/ISRIC/ISS-CAS/JRC[71]
Slope	1. slopedeg 2. slopepct	1. Slope in degree generated from the elevation 2. Slope in percent generated from the elevation	raster	FAO/IIASA/ISRIC/ISS-CAS/JRC[71]
Land Cover	lc2010	Land-cover in 2010 [30 m resolution]	raster	CCRS/CCMEO/NRCan/CONABIO/CONAFOR/INEGI/USGS[72]
Land Use	1. cdl2008, cdl2009, cdl2010, cdl2011, cdl2012, cdl2013, cdl2014, cdl2015, cdl2016, cdl2017, cdl2018 2. MX_LU_2011, MX_LU_2014 3. MX_ancrop_winter07, MX_ancrop_spring07, MX_percrop07	1. Time-series (2008–2018) of the Cropland Data Layer in US [30 m resolution] 2. Land-Use in MX in 2011 and 2014 [30 m resolution] 3. Annual Winter, Annual Spring and Perennial Areas Planted in MX in 2007 [*municipio*-based]	raster, polygon	USDA-NASS[73], INEGI[74,75], INEGI[42]

Table 4. List of provided geospatial layers for the Biophysical Environment category including their description, format and input data sources.

Wildlife Service recently discontinued its funding for the LCCs network, we decided to include these cooperatives in our dataset since the LCCs generated highly relevant information for the RGB in the past. The third project refers to the Instream Flow Program in Colorado. For this project, we mapped the streams and lakes affected by the Instream Flow Program.

Hydrology (HYD). The Hydrology category includes five surface water and groundwater related sub-categories.

Watershed boundary. The geodatabase provides the outlines of the Rio Grande/Río Bravo basin and of its sub-basins. The watershed boundary dataset is derived from the World Wildlife Fund's (WWF) HydroSHEDS product, created from NASA's Shuttle Radar Topographic Mission (SRTM) 15-second Digital Elevation Model (DEM) and released by the Food and Agriculture Organization (FAO) geoportal.

River networks. This dataset of vectorized river reaches, derived from the WWF HydroSHEDS products, was also accessed from the FAO geoportal.

Aquifer boundaries. We mapped the aquifer boundaries by assembling two national geospatial datasets compiled by the National System of Water Information of CONAGUA and the USGS.

Gauging stations. The geodatabase provides datasets for the location of discharge monitoring stations operated by USGS (above Elephant Butte Dam), IBWC/CILA (below Elephant Butte Dam and along the border to the Gulf of Mexico), and CONAGUA (tributaries of the Rio Grande/Río Bravo coming from Mexico). Each dataset provides a URL to the historical time series of the stream flow.

Water quality stations. This dataset provides the location of water quality stations monitored by IBWC and TCEQ for the U.S. portion of the RGB.

Water Use & Hydraulic Infrastructures (USE). Water Use & Hydraulic Infrastructures datasets provide information related to consumption and diversion of water.

Withdrawals. This dataset provides estimates in million cubic meters per year of water used at the U.S. county and MX *municipio* level. For both countries, estimates are available by type of source (surface or groundwater) and by sector of use (e.g., public supply, domestic, industrial, irrigation, thermoelectricity). For the U.S., the data set also breaks down estimates by fresh water and saline water.

Irrigation. We provide two datasets (one per country), which reports on a county/*municipio* basis the irrigated area estimated in hectares and the spatial coverage of different systems of irrigation (surface, sprinkler, micro-irrigation and other), estimated in hectares for U.S. and in number of farms for Mexico. For the U.S., we used the original classes (surface, sprinkler, micro-irrigation) and for Mexico, we clustered as follows: lined and earthen canals in surface; sprinkler and micro-sprinkler in sprinkler; and drip in micro-irrigation.
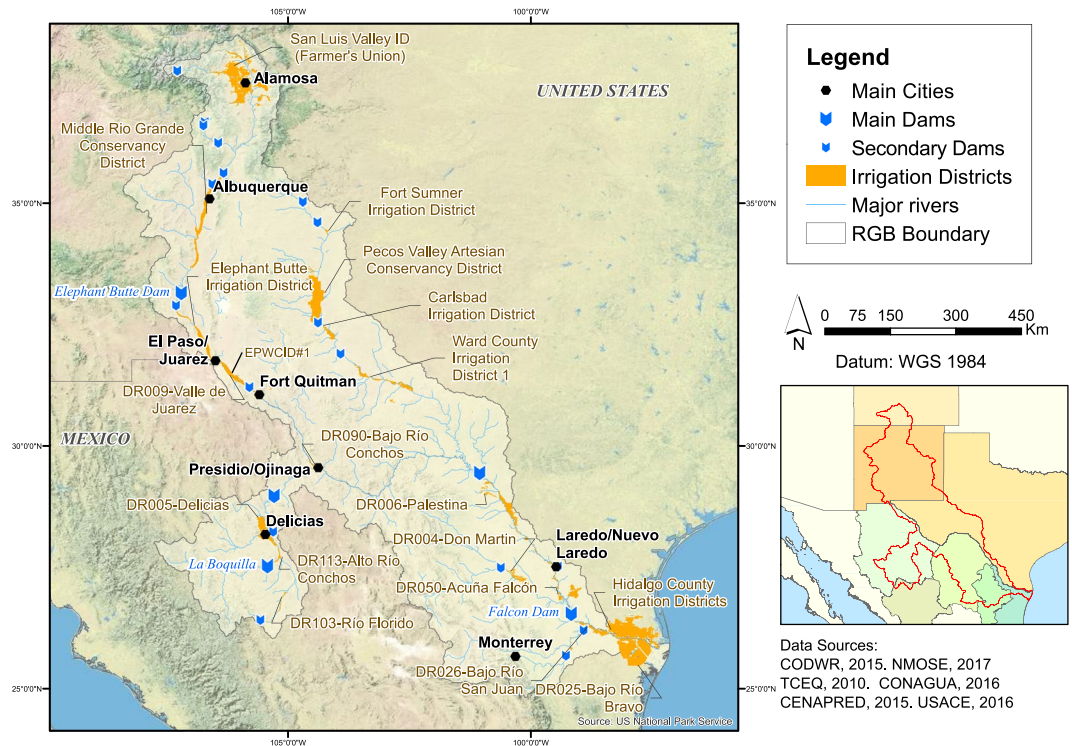
Fig. 2 Map of the main dams and irrigation districts in the Rio Grande/Río Bravo basin.

Water rights/permits. This sub-category displays the allocation of surface water rights and groundwater rights at the state level for U.S. and Mexico. A water right is a property right that conveys the right to use a certain amount of water. These are administered by the Offices of the State Engineer in the U.S. and by CONAGUA in Mexico. The datasets report the geographic coordinates of the water right holder, and when available, the amount of water right authorized (in acre feet in the U.S. and in cubic meters in Mexico), the name of the water right holder, the adjudication date, the rank of priority (i.e., the water right's seniority in a water drainage), or the type of use.

Dams. We provide two vector datasets. The first one, derived from the National Inventory of Dams, maps all dams in the U.S. exceeding a certain height and capacity of storage. The second one maps all dams recorded by the CENAPRED in Mexico.

Water diversion structures. The geodatabase provides three vector datasets of water diversion structures (ditch/canal, pipeline, underground conduit and connector). The first one covers the five MX states and is derived from a Mexican national data source. The second one covers the three U.S. states and is derived from a U.S. national data source. The third one covers Colorado with more details and is derived from a state data source.

Wells (included only for the U.S.). We compiled four datasets of well locations. One is derived from the National Hydrography Dataset (NHD) for the whole U.S. portion of the RGB; the three other ones were compiled by the OSE for each U.S. state and provides more details. Our research to-date has not uncovered data for wells in Mexico.

Socio-Economics (SE). This category includes five datasets related to demography, economics, and transportation networks.

Population. This vector dataset reports the total population for the years 1990, 2000, and 2010 on a county/*municipio* basis. We assembled national demographic census information distributed by the U.S. Census Bureau and INEGI.

Population density. This raster dataset depicts population density (number of persons per square kilometre) at 30 arc-second resolution (approximately 1 km at the Equator) for the years 2000, 2005, 2010, 2015, and 2020.

Income. Due to the lack of consistent data for both countries, the spatial database includes two heterogeneous datasets: the personal income for the U.S. counties from 1969 to 2015 and the income distribution (i.e., the distribution of the working population per category of income) in 1990, 2000, and 2010 for the Mexican *municipios*.
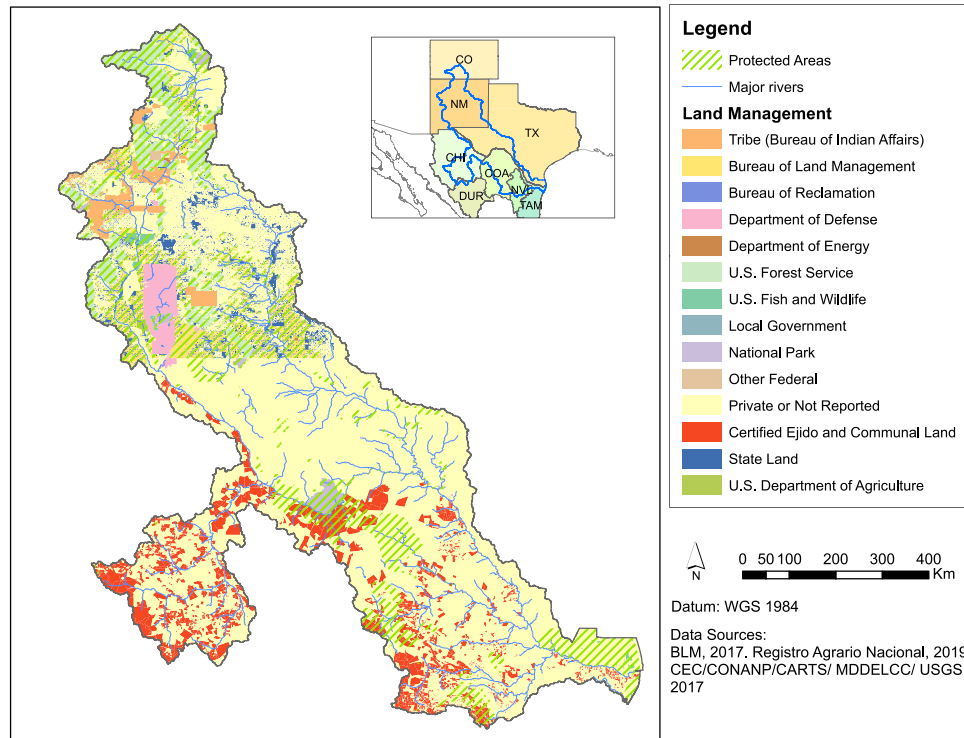
Fig. 3 Map of the Rio Grande/Río Bravo basin depicting land ownership, protected areas, and major rivers, generated from the geospatial datasets of the geodatabase.

Number and size of farms.   This sub-category provides county/*municipio*-based statistics about the number of farms, the total farmed area, and the average farm size (in 2007 and 2012 for the U.S. and in 2007 for Mexico).

Transport infrastructures.   This sub-category contains two datasets: the roads derived from the Global Roads Open Access Data Set, Version 1 (gROADSv1) and the railroads derived from the North American Environmental Atlas.

Biophysical environment (BIO).   The biophysical environment category provides ecological, pedological, topographical, and land-use/land-cover datasets.

Ecoregions.   Ecoregions are areas with similar patterns of biotic and abiotic phenomena, including geology, physiography, vegetation, climate, soils, land use, wildlife, and hydrology. Ecoregions can be mapped for different nested levels. For North America (Canada, U.S. and Mexico), data are available for three levels (Level I being the coarsest and Level III the finest). Ecoregions level IV have also been delineated for the U.S. but not yet for Mexico. Each of those datasets are included in the geodatabase, although ecoregion Level IV is missing for Mexico.

Habitat.   We compiled six national datasets due to a lack of transboundary coverage. In the U.S., we mapped wetlands and deepwater habitats (extent, approximate location, and type), riparian areas, and critical habitat for endangered and threatened species. In Mexico, we mapped Ramsar wetlands sites (representative, rare or unique wetlands that are important to conserve biological diversity), priority sites for restoration (areas of high biological value that require restoration actions), and priority focus sites for biodiversity conservation (conserved habitats that host endangered and threatened species and are adjacent to the protected areas).

Soil cover.   This vector dataset displays different soil units based on the FAO taxonomy at 1:5,000,000 scale.

Elevation.   This raster data, derived from the Harmonized World Soil Database (version 1.1), provides the median elevation aggregated to 30 arc-second grid cells. It is derived from the NASA Shuttle Radar Topographic Mission Digital Elevation Model and the USGS GTOPO30.

Slope.   We calculated the median slope in percent and in degree at a 30 arc-second resolution from the elevation data described above, by using the *Spatial Analyst* tool Slope in ArcGIS.

Land cover.   This raster dataset, derived from the North American Land Change Monitoring System (NALCMS), depicts the land cover in the RGB, based on 2010 Landsat satellite imagery. NALCMS uses the level II of the Land Cover Classification System (LCCS) standard developed by FAO.
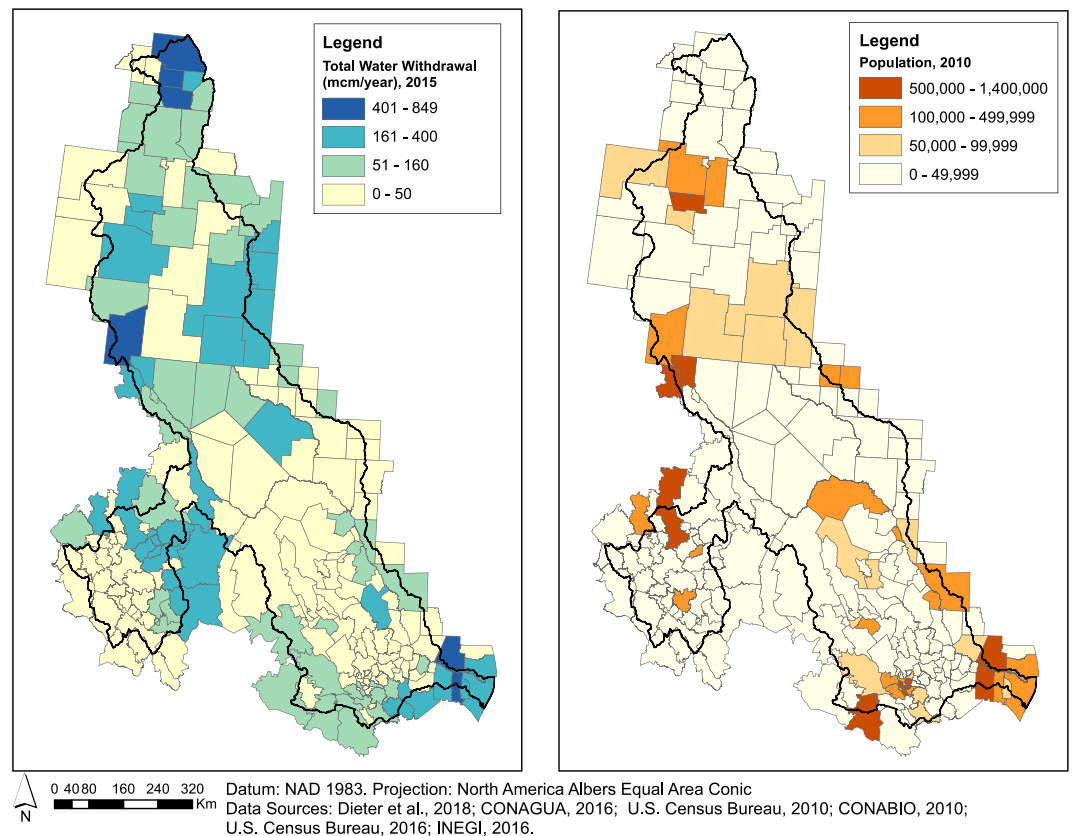
0 40 80   160   240   320
                          Km

Datum: NAD 1983. Projection: North America Albers Equal Area Conic
Data Sources: Dieter et al., 2018; CONAGUA, 2016; U.S. Census Bureau, 2010; CONABIO, 2010;
U.S. Census Bureau, 2016; INEGI, 2016.

Fig. 4 Map comparing the total amount of water withdrawals (millions of cubic meters per year) in 2015 with the 2010 population on a county/*municipio* basis.

Land use.   Due to a lack of consistent transboundary datasets, we gathered national datasets to map land-use in the RGB. For the U.S., we used the yearly USDA Cropland Data Layer, a high-resolution crop-specific land-use image, available from 2008 to 2018 at a 30-meter resolution. For Mexico, we derived our datasets from two governmental datasets. The first one is a spatial dataset of land uses for 2011 and 2014 at a 30-meter resolution, with a coarse classification of crops (rain-fed *vs* irrigated crops). The second one, non-spatial, is derived from the 2007 Census and provides on a *municipio* basis planted area estimates (in hectares) of annual spring and winter crops and perennial crops.

Data gaps.   Although we developed an extensive database, it is not comprehensive for either the U.S. or Mexico. Our approach led us to identify several spatial data gaps for socio-environmental research in the RGB.

Yearly crop-specific datasets for Mexico.   Supplying yearly crop-specific time-series datasets at a 30-m resolution will be helpful for comparing cropland dynamics in the U.S. and Mexico.

Boundaries of community ditch and acequia associations (in NM) and Unidades de Riego (MX).   To the authors' knowledge, these spatial boundaries are not publicly available. Mapping them is important, because of their water management implications. The NMOSE is currently working on updating and assembling this information for New Mexico[31].

Land ownership.   For Mexico, we did not find any spatial datasets delineating public lands, except the Certified Ejidos and Communal Lands. Generating a dataset that differentiates, in more detail, public from private lands in Mexico would be valuable for land and water management research.

Groundwater wells (MX).   We were not able to access a spatial dataset of wells for Mexico. Mapping well locations is of primary interest for surface and groundwater management in the basin.

Binational infrastructures.   Some information on binational infrastructure is missing, such as the Morillo drain (described in Supplementary Information 1).

Climate records are also critical for understanding hydrological processes in the basin. While not included in our geodatabase, Livneh and others[11] published a transboundary historic dataset that derives observed daily precipitation, minimum and maximum temperature gridded to a 1/16° (~6 km) resolution for the time period 1950–2013.

Technical Validation

To ensure the reliability of the spatial database, we draw on a critical assessment of three factors: the quality of the source of information, transboundary discontinuities, and geoprocessing.

Regarding the sources of information, we sought to download most of the raw datasets from official sources such as intergovernmental organizations and governmental agencies (e.g., FAO, USIBWC, USGS, USDA, CONAGUA, INEGI, CONABIO, CONANP, CODWR, NMOSE, TCEQ) or from joint projects overseen by reputable universities, research institutes, and non-governmental organizations (e.g., from the NASA Socioeconomic Data and Applications Center, the North American Environmental Atlas). We compiled a few datasets (around 6%) from private entities (associations, districts) as they were the only source of the desired information. For example, this is the case for the border fences. We downloaded the spatial data from Reveal from The Center for Investigative Reporting and OpenStreetMap contributors (https://github.com/cirlabs/border_fence_map) who traced the fences along the border by using open-source mapping tools and by digitizing detailed PDF maps obtained through a Freedom of Information Act request[32]. The creators of the data set acknowledge slight differences with the official government agency numbers. However, this is the most detailed border fence map publicly available. We also gathered four datasets — that were not downloadable online — by email requests from three organizations. The RGWCD provided the boundaries of the RGWCD and the Groundwater Management Subdistricts #1, #2, #3, #4, #5, #6 in Colorado. The CSCB (Colorado State Conservation Board) and the TSSWCB (Texas State Soil and Water Conservation Board) shared the boundaries of the SWCDs of Colorado and Texas, respectively. Although these datasets were not gathered from governmental sources, we believe that these datasets are the best available and we rely on the accuracy of data collection of the three organizations.

Whenever possible, we used global or bi-national input datasets to avoid inconsistencies related to differences in resolution, type of information, or geoprocessing methods among counties and states. Around 11.5% of our raw datasets have a global or bi-national spatial extent, and therefore cross international and state boundaries. Nevertheless, a significant amount of the raw datasets we gathered were nation-wide (70%) or state-wide (18.5%), which could lead to some of the inconsistencies previously mentioned. All information gathered from Mexican sources was translated from Spanish to English and verified by a bilingual speaker.

Regarding the geoprocessing operations, for all datasets, we used several tools and functions available on ArcGIS, such as Project, Clip, Dissolve, Slope, Join, Merge and we rely on the accuracy of this software to create the final outputs. A series of quality checks has been carried out to catch as many potential coding errors as possible. Attributes were checked by using visual inspection. For quality control, transparency, and reproducibility, processing steps are thoroughly detailed in the metadata of each dataset.

Usage Notes

This spatial database offers a wide range of applications for transboundary and interdisciplinary research. We provide three examples drawing from our own research.

1. A broad range of mapping products can be created for the whole RGB or sub-sections, such as a map of the main irrigation districts and dams in the RGB (Fig. 2), a map of land managers and protected areas (Fig. 3), and a comparative map of water use and population on a county/*municipio* basis (Fig. 4).
2. The geodatabase can support spatially explicit and integrated modelling for the RGB[27]. We are currently developing a simulation model for the RGB that draws on water and land governance datasets (political jurisdiction boundaries, land management and protected areas, irrigation district outlines), as well as hydrologic and biophysical environment datasets (river network, elevation, water bodies, gauges, land-use), socio-economic (population), and water use and hydraulic infrastructures (dams).
3. The geodatabase can be used to conduct spatiotemporal and geostatistical analyses. Examples of interdisciplinary topics include: (i) evaluating whether water use correlates with environmental variables (ecoregions, land use), socio-economic factors (population), or governance (e.g., related to land ownership, irrigation district, or state); and (ii) assessing land-use change at multiple scales (local, state, national). For example, we are currently analysing the spatial patterns of land fallowing in the basin.

Users of the compiled datasets should cite this data paper following the recommended citation format of the journal. Furthermore, we encourage users to refer to the sections "Credits" and "Use limitations" in the metadata of each layer file, as additional data constraints and required credits of the original data sources may apply.

In conclusion, to achieve more sustainable transboundary water management, it is necessary to approach problems from a boundary-spanning and interdisciplinary perspective. By creating a basin-wide, socio-environmental geodatabase for the RGB, we address this issue and support further interdisciplinary research in the basin. The main novelty of our approach was to include geospatial datasets mapping public and private water governance, which contributes to better understanding of scale interactions and mismatches between ecological and social (decision-making) boundaries[33]. Moreover, we were able to identify data gaps, especially for the Mexican side of the RGB, and therefore the need for further research to reduce data imbalance between the two sides of this transboundary basin. Even though our study is limited to the socio-environmental boundaries of the RGB system and the typology of actors is specific to the shared basins between U.S. and Mexico, we consider that our approach to assembling a wide range of data is transferable to other shared river basins to serve transboundary and cross-disciplinary collaboration and research.

Code availability

We created the individual output datasets using geoprocessing scripts in the Python programming language and the arcpy package. All geoprocessing scripts are available on GitHub and are accessible through the Open Science Framework (OSF) repository[28].

References

1. Garcia, X. & Pargament, D. Reusing wastewater to cope with water scarcity: Economic, social and environmental considerations for decision-making. *Resour. Conserv. Recy.* **101**, 154–166 (2015).
2. Green, P. A. *et al.* Freshwater ecosystem services supporting humans: Pivoting from water crisis to water solutions. *Glob. Environ. Chang.* **34**, 108–118 (2015).
3. Jaeger, W. K. *et al.* Finding water scarcity amid abundance using human–natural system models. *Proc. Nat. Acad. Sci. USA* **114**, 11884–11889 (2017).
4. WWAP (World Water Assessment Programme). *Managing Water Under Uncertainty and Risk*. The United Nations World Water Development Report 4. UNESCO. Paris, France (2012).
5. United Nations. *The Sustainable Development Goals Report 2019*. New York, NY (2019).
6. Munia, H. *et al.* Water stress in global transboundary river basins: significance of upstream water use on downstream stress. *Environ. Res. Lett.* **11** (2016).
7. UNECE/UNESCO. *Good Practices in Transboundary Water Cooperation* (2015).
8. Troy, T. J., Konar, M., Srinivasan, V. & Thompson, S. Moving sociohydrology forward: A synthesis across studies. *Hydrol. Earth Syst. Sci.* **19**, 3667–3679 (2015).
9. Leslie, H. M. *et al.* Operationalizing the social-ecological systems framework to assess sustainability. *Proc. Nat. Acad. Sci. USA* **112**, 5979–5984 (2015).
10. Hanspach, J. *et al.* A holistic approach to studying social-ecological systems and its application to southern Transylvania. *Ecol. Soc.* **19** (2014).
11. Livneh, B. *et al.* A spatially comprehensive, hydrometeorological data set for Mexico, the U.S., and Southern Canada 1950–2013. *Sci. Data.* **2**, 1–12 (2015).
12. Tucker Lima, J. M. *et al.* A social-ecological database to advance research on infrastructure development impacts in the Brazilian Amazon. *Sci. Data.* **3**, 1–9 (2016).
13. UNEP & Oregon State University. *Atlas of International Freshwater Agreements*. Vol. 4 (United Nations Environmental Programme, 2002).
14. Nava, L. F. & Sandoval-Solis, S. Multi-tiered governance of the Rio Grande/Bravo Basin: The fragmented water resources management model of the United States and Mexico. *International Journal of Water Governance.* **2**, 85–106 (2014).
15. Dean, D. J. & Schmidt, J. C. The role of feedback mechanisms in historic channel changes of the lower Rio Grande in the Big Bend region. *Geomorphology.* **126**, 333–349 (2011).
16. Blythe, T. L. & Schmidt, J. C. Estimating the natural flow regime of rivers with long-standing development: The Northern Branch of the Rio Grande. *Water Resour. Res.* **54**, 1212–1236 (2018).
17. Wong, C. M., Pittock, J. & Schelle, P. *World's Top 10 Rivers at Risk*. Gland, Switzerland (WWF International, 2007).
18. Sandoval-Solis, S., Teasley, R. L., McKinney, D. C., Thomas, G. A. & Patiño-Gomez, C. Collaborative modeling to evaluate water management scenarios in the Rio Grande Basin. *J. Am. Water Resour. Assoc.* **49**, 639–653 (2013).
19. Comisión Nacional del Agua (CONAGUA). *Statistics on Water in Mexico, 2010 Edition*. Ministry of Environment and Natural Resources. México, D.F. (2010).
20. Maupin, M. A. *et al. Estimated use of water in the United States in 2010*. U.S. Geological Survey Circular 1405, https://doi.org/10.3133/cir1405, (U.S. Geological Survey, 2014).
21. Jun, C., Ban, Y. & Li, S. China: Open access to Earth land-cover map. *Nature.* **514**, 434–434 (2014).
22. Mu, J. E. & Ziolkowska, J. R. An integrated approach to project environmental sustainability under future climate variability: An application to U.S. Rio Grande Basin. *Ecol. Indic.* **95**, 654–662 (2018).
23. Llewellyn, D. & Vaddey, S. *West-Wide Climate Risk Assessment: Upper Rio Grande Impact Assessment* (U.S. Bureau of Reclamation, 2013).
24. Wolf, A. T. The transboundary freshwater dispute database project. *Water Int.* **24**, 160–163 (1999).
25. Ortiz-Partida, J. P., Sandoval-Solis, S. & Diaz-Gomez, R. Assessing the State of Water Resource Knowledge and Tools for Future Planning in the Rio Grande-Rio Bravo Basin, https://doi.org/10.21429/C9BC7D (2017).
26. Patiño-Gomez, C., McKinney, D. C. & Maidment, D. R. Sharing water resources data in the binational Rio Grande/Bravo Basin. *J. Water Resour. Plan. Manag.* **133**, 416–426 (2007).
27. Koch, J., Friedman, J. R., Paladino, S., Plassin, S. & Spencer, K. Conceptual modeling for improved understanding of the Rio Grande/Río Bravo socio-environmental systems. *Socio-Environmental Systems Modelling.* **1** (2019).
28. Plassin, S. *et al.* Geospatial data for the Rio Grande/Río Bravo socio-environmental system. *Open Science Framework.* https://doi.org/10.17605/OSF.IO/79426 (2019).
29. DiNatale Water Consultants. *Rio Grande Basin Implementation Plan*. Alamosa, CO (Rio Grande Basin Roundtable, 2015).
30. U.S. Geological Survey (USGS). *The Transboundary Aquifer Assessment Program (TAAP)*, https://webapps.usgs.gov/taap/.
31. New Mexico Office of the State Engineer (NMOSE), DHSEM, EDAC & FEMA. *Acequia Mapping Project outreach*, https://ose.maps.arcgis.com/apps/Cascade/index.html?appid=b6f0edf3d28a49dd822c558658b9a35d.
32. Corey, M. & Becker, A. *The wall: Building a continuous US-Mexico barrier would be a tall order*, https://www.revealnews.org/article/the-wall-building-a-continuous-u-s-mexico-barrier-would-be-a-tall-order/ (2017).
33. Cumming, G., Cumming, D. H. M. & Redman, C. Scale mismatches in social-ecological systems: causes, consequences, and solutions. *Ecol. Soc.* **11** (2006).
34. Food and Agriculture Organization (FAO). Hydrological basins in Central America (Derived from HydroSHEDS), http://www.fao.org/geonetwork/srv/en/main.home (2009).
35. Food and Agriculture Organization (FAO). Rivers in Central America (Derived from HydroSHEDS), http://www.fao.org/geonetwork/srv/en/main.home (2014).
36. U.S. Geological Survey (USGS). Ground Water Atlas of the United States. Aquifers vector digital data, https://catalog.data.gov/dataset/aquifers (2008).
37. Comisión Nacional del Agua (CONAGUA). Sistema Nacional de Información del Agua (SINA), http://201.116.60.25/sina/ (2018).
38. Stewart, D. W., Rea, A. & Wolock, D. M. USGS Streamgages Linked to the Medium Resolution NHD, https://doi.org/10.3133/ds195 (2006).
39. U.S. International Boundary and Water Commissions (IBWC). *GIS Portal*, https://appportal.ibwc.gov/ibwc_geo/public_portal/ (2018).

40. Comisión Nacional del Agua (CONAGUA). Base Datos Bandas, ftp://ftp.conagua.gob.mx/Bandas/Bases_Datos_Bandas (2017).
41. Dieter, C. A. *et al.* Estimated Use of Water in the United States County-Level Data for 2015 (ver. 2.0, June 2018). U.S. Geological Survey data release, https://doi.org/10.5066/F7TB15V5 (2018).
42. Instituto Nacional de Estadística y Geografía (INEGI). Censo Agrícola, Ganadero y Forestal 2007, https://www.inegi.org.mx/programas/cagf/2007/default.html (2009).
43. Colorado Division of Water Resources (CDWR). DWR Water Right - Net Amounts, https://data.colorado.gov/Water/DWR-Water-Right-Net-Amounts/acsg-f33s/data (2019).
44. New Mexico Office of the State Engineer (NMOSE). *Open Data Site*, http://geospatialdata-ose.opendata.arcgis.com/ (2018).
45. Texas Commission on Environmental Quality (TCEQ). WRAP Input files and GIS files by River Basin, https://www.tceq.texas.gov/permitting/water_rights/wr_technical-resources/wam.html/#wrapinput (2013).
46. Comisión Nacional del Agua (CONAGUA). Tablero SINA: Registro Público de Derechos de Agua (Repda)/Volúmenes inscritos, http://siga.conagua.gob.mx/REPDA/Menu/MenuKMZ.html (2017).
47. U.S. Army of Corps Engineer (USACE). 2018 National Inventory of Dams (NID), http://nid.usace.army.mil/ (2019).
48. Centro Nacional de Prevención de Desastres (CENAPRED). Presas, http://catalogo.datos.gob.mx/dataset/presas (2015).
49. U.S. Geological Survey (USGS). National Hydrography Dataset – Medium resolution (1:100,000), https://www.usgs.gov/core-science-systems/ngp/national-hydrography (2014).
50. Instituto Nacional de Estadística y Geografía (INEGI). Conjunto de datos vectoriales de la serie topográfica escala 1:1,000,000. Acueducto y Canal, https://www.inegi.org.mx/app/biblioteca/ficha.html?upc=702825267643 (2000).
51. Colorado Water Conservation Board (CWCB) & Division of Water Resources (DWR). *Colorados' Decision Support Systems*, https://www.colorado.gov/pacific/cdss/gis-data-category (2014).
52. Texas Water Development Board (TWDB). *GIS Datasets*, http://www.twdb.texas.gov/mapping/gisdata.asp (2018).
53. U.S. Census Bureau. Profile of General Population and Housing Characteristics: 1990, 2000, 2010, https://factfinder.census.gov/faces/nav/jsf/pages/searchresults.xhtml?refresh=t (2010).
54. Instituto Nacional de Estadística y Geografía (INEGI). Censos y Conteos de Población y Vivienda. Serie histórica censal e intercensal (1990–2010), https://www.inegi.org.mx/programas/ccpv/cpvsh/ (2018).
55. Center for International Earth Science Information Network (CIESIN) - Columbia University. Gridded Population of the World, Version 4 (GPWv4): Population Density, Revision 11, https://doi.org/10.7927/H49C6VHW (2018).
56. U.S. Bureau of Economic Analysis. Annual Personal Income by State: 1969–2015, https://apps.bea.gov/regional/Downloadzip.cfm (2016).
57. Comisión Nacional para el Conocimiento y Uso de la Biodiversidad (CONABIO). Ingreso en México por municipio, 1990, 2000, 2010. Datos estadísticos del Instituto Nacional de Estadísitca y Geografía (INEGI), http://www.conabio.gob.mx/informacion/gis/ (2012).
58. U.S. Department of Agriculture - National Agricultural Statistics Service (USDA-NASS). Census of Agriculture 2012 and 2007, https://www.nass.usda.gov/AgCensus/index.php (2014).
59. Center for International Earth Science Information Network (CIESIN) - Columbia University & Information Technology Outreach Services (ITOS) - University of Georgia. Global Roads Open Access Data Set, Version 1 (gROADSv1), https://doi.org/10.7927/H4VD6WCT (2013).
60. Instituto Nacional de Estadística y Geografía (INEGI), Natural Resources Canada (NRCan), U.S. Geological Survey (USGS) & Commission for Environmental Cooperation (CEC). North American Environmental Atlas - Railroads, 2009, http://www.cec.org/tools-and-resources/north-american-environmental-atlas/map-files (2010).
61. U.S. Environmental Protection Agency (EPA). Level III and IV ecoregions of the continental United States, https://www.epa.gov/eco-research/level-iii-and-iv-ecoregions-continental-united-states (2013).
62. U.S. Environmental Protection Agency (EPA). Level III ecoregions of North America, https://www.epa.gov/eco-research/ecoregions-north-america (2010).
63. U.S. Fish & Wildlife Service (USFWS). National Wetlands Inventory. A System for Mapping Riparian Areas in the Western United States, https://www.fws.gov/wetlands/Other/Riparian-Product-Summary.html (2018).
64. U.S. Fish & Wildlife Service (USFWS). National Wetlands Inventory - Version 2 - Surface Waters and Wetlands Inventory, https://www.fws.gov/wetlands/Data/Data-Download.html (2018).
65. U.S. Fish & Wildlife Service (USFWS). Threatened & Endangered Species Active Critical Habitat Report, https://ecos.fws.gov/ecp/report/table/critical-habitat.html (2019).
66. Comisión Nacional de Áreas Naturales Protegidas (CONANP). Sitios RAMSAR, https://datos.gob.mx/busca/dataset/coberturas-para-manejadores-de-sig (2015).
67. Comisión Nacional para el Conocimiento y Uso de la Biodiversida (CONABIO). Sitios de atención prioritaria para la conservación de la biodiversidad, http://www.conabio.gob.mx/informacion/gis/ (2016).
68. Comisión Nacional para el Conocimiento y Uso de la Biodiversida (CONABIO). Sitios prioritarios para la restauración, http://www.conabio.gob.mx/informacion/gis/ (2016).
69. Food and Agriculture Organization (FAO). The Digital Soil Map of the World, Version 3.6, http://www.fao.org/geonetwork/srv/en/metadata.show?id=14116 (2003).
70. Sanchez, P. A. *et al.* Digital Soil Map of the World. *Science.* **325**, 680 (2009).
71. FAO/IIASA/ISRIC/ISS-CAS/JRC. Harmonized World Soil Database (version 1.1), http://webarchive.iiasa.ac.at/Research/LUC/Products-Datasets/global-terrain-slope-download.html (2009).
72. Canada Centre for Remote Sensing (CCRS)/Canada Centre for Mapping and Earth Observation (CCMEO) Natural Resources Canada (NRCan), Comisión Nacional para el Conocimiento y Uso de la Biodiversidad (CONABIO), Comisión Nacional Forestal (CONAFOR), Insituto Nacional de Estadística y Geografía (INEGI) & U.S. Geological Survey (USGS). 2010 Land Cover of North America at 30 meters. Edition 1.0, http://www.cec.org/tools-and-resources/north-american-environmental-atlas/map-files (2017).
73. U.S. Department of Agriculture - National Agricultural Statistics Service (USDA-NASS). 2008–2018 Cropland Data Layer, https://nassgeodata.gmu.edu/CropScape/ (2019).
74. Instituto Nacional de Estadística y Geografía (INEGI). Conjunto de datos vectoriales de Uso del Suelo y Vegetación 2011, Escala 1:250,000, Serie V (Capa Unión), https://www.inegi.org.mx/temas/usosuelo/ (2013).
75. Instituto Nacional de Estadística y Geografía (INEGI). Conjunto de datos vectoriales de Uso del Suelo y Vegetación 2014, Escala 1:250,000, Serie VI (Capa Unión), https://www.inegi.org.mx/temas/usosuelo/ (2016).
76. U.S. Census Bureau. 2016 TIGER/Line**®** Shapefiles, https://www.census.gov/ (2016).
77. Instituto Nacional de Estadística y Geografía (INEGI). Marco Geoestadístico, http://www.beta.inegi.org.mx/app/biblioteca/ficha.html?upc=702825217341 (2016).
78. Instituto Nacional de Estadística y Geografía (INEGI), Natural Resources Canada (NRCan), U.S. Geological Survey (USGS) & Commission for Environmental Cooperation (CEC). North American Environmental Atlas - Populated Places, 2009, http://www.cec.org/tools-and-resources/north-american-environmental-atlas/map-files (2010).
79. Texas Commission on Environmental Quality (TCEQ). *GIS Data*, https://www.tceq.texas.gov/gis/download-tceq-gis-data (2018).
80. U.S. Geological Survey & U.S. Department of Agriculture Natural Resources Conservation Service. *Federal Standards and Procedures for the National Watershed Boundary Dataset (WBD) (4 ed.).* Techniques and Methods 11–A3 (U.S. Geological Survey, 2013).

81. Comisión Nacional del Agua (CONAGUA). Regiones Hidrológicas Administrativas. (Organismos de Cuenca). Shapefile. Escala: 1:250000, http://www.conabio.gob.mx/informacion/gis/ (2007).
82. Rio Grande Water Conservation District (RGWCD). Well information, https://www.rgwcd.org/well-information.
83. Teeple, A. P. *Geophysics- and geochemistry-based assessment of the geochemical characteristics and groundwater-flow system of the U.S. part of the Mesilla Basin/Conejos-Médanos aquifer system in Doña Ana County, New Mexico, and El Paso County, Texas, 2010–12*. Scientific Investigations Report. Reston, VA, https://doi.org/10.3133/sir20175028, (U. S. Geological Survey, 2017).
84. Driscoll, J. M. & Sherson, L. R. *Variability of surface-water quantity and quality and shallow groundwater levels and quality within the Rio Grande Project Area, New Mexico and Texas, 2009–13*. Scientific Investigations Report. Reston, VA, https://doi.org/10.3133/sir20165006, (U. S. Geological Survey, 2016).
85. Bureau of Land Management (BLM). National Surface Management Agency Area Polygons - National Geospatial Data Asset, https://landscape.blm.gov/geoportal/catalog/main/portal.page (2017).
86. Registro Agrario Nacional (RAN). Perimetrales de los núcleos agrarios certificados, http://datos.ran.gob.mx/conjuntoDatosPublico.php (2019).
87. Commission for Environmental Cooperation (CEC), Comisión Nacional de Áreas Naturales Protegidas (CONANP), Conservation Areas Reporting and Tracking System (CARTS), Ministère du developpement durable et de la lutte contre le changement climatique (Quebec-MDDELCC) & U.S. Geological Survey (USGS). Protected Areas of North America, 2017, http://www.cec.org/tools-and-resources/north-american-environmental-atlas/map-files (2017).
88. Homeland Infrastructure Foundation-Level Data (HIFLD). Canada and Mexico Border Crossings, https://hifld-geoplatform.opendata.arcgis.com/datasets/canada-and-mexico-border-crossings (2010).
89. Homeland Security Infrastructure Program (HSIP). Freedom Office of Border Patrol Sectors, https://www.arcgis.com/home/item.html?id=e4c86699f8b84b6f9e5b26bf452323a8 (2017).
90. U.S. Customs and Border Protection (USCPB). Southwest Border Migration Apprehension Statistics. Fiscal years 2017 and 2018, https://www.cbp.gov/newsroom/stats/usbp-sw-border-apprehensions (2019).
91. U.S. Fish & Wildlife Service (USFWS). 2015 LCC Network Areas, https://www.sciencebase.gov/catalog/item/55b943ade4b09a3b01b65d78 (2015).
92. U.S. Fish & Wildlife Service (USFWS). North American Joint Ventures, https://ecos.fws.gov/ServCat/Reference/Profile/81433 (2017).

Acknowledgements

Author contributions

J.K. and J.R.F. designed the study. S.Pa. and J.R.F. conducted the ethnographic research, and proposed the qualitative typology of actors. All co-authors contributed in the identification of datasets. S.Pl. collected and processed the input datasets, generated the spatial database and created the outputs with contributions from J.K., K.S. and K.V. S.Pa. verified and edited Spanish-English translation. S.Pl. led the manuscript writing and all authors reviewed and edited the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41597-020-0410-1.

Correspondence and requests for materials should be addressed to J.K.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.