



Democratizing Glacier Data – Maturity of Worldwide Datasets and Future Ambitions

Isabelle Gärtner-Roer^{1*}, Samuel U. Nussbaumer¹, Bruce Raup², Frank Paul¹, Ethan Welty¹, Ann K. Windnagel², Florence Fetterer² and Michael Zemp¹

¹ World Glacier Monitoring Service (WGMS), Department of Geography, University of Zurich, Zurich, Switzerland, ² National Snow and Ice Data Center (NSIDC), Cooperative Institute for Research in Environmental Sciences (CIRES), University of Colorado Boulder, Boulder, CO, United States

OPEN ACCESS

Edited by:

Tiffany C. Vance,
U.S. Integrated Ocean Observing
System, United States

Reviewed by:

Wesley Van Wychen,
University of Waterloo, Canada
Laura Thomson,
Queen's University, Canada

*Correspondence:

Isabelle Gärtner-Roer
isabelle.roer@geo.uzh.ch

Specialty section:

This article was submitted to
Climate Services,
a section of the journal
Frontiers in Climate

Received: 21 December 2021

Accepted: 17 May 2022

Published: 27 June 2022

Citation:

Gärtner-Roer I, Nussbaumer SU,
Raup B, Paul F, Welty E,
Windnagel AK, Fetterer F and Zemp M
(2022) Democratizing Glacier Data –
Maturity of Worldwide Datasets and
Future Ambitions.
Front. Clim. 4:841103.
doi: 10.3389/fclim.2022.841103

The creation and curation of environmental data present numerous challenges and rewards. In this study, we reflect on the increasing amount of freely available glacier data (inventories and changes), as well as on related demands by data providers, data users, and data repositories in-between. The amount of glacier data has increased significantly over the last two decades as remote sensing techniques have improved and free data access is much more common. The portfolio of observed parameters has increased as well, which presents new challenges for international data centers, and fosters new expectations from users. We focus here on the service of the Global Terrestrial Network for Glaciers (GTN-G) as the central organization for standardized data on glacier distribution and change. Within GTN-G, different glacier datasets are consolidated under one umbrella, and the glaciological community supports this service by actively contributing their datasets and by providing strategic guidance via an Advisory Board. To assess each GTN-G dataset, we present a maturity matrix and summarize achievements, challenges, and ambitions. The challenges and ambitions in the democratization of glacier data are discussed in more detail, as they are key to providing an even better service for glacier data in the future. Most challenges can only be overcome in a financially secure setting for data services and with the help of international standardization as, for example, provided by the CoreTrustSeal. Therefore, dedicated financial support for and organizational long-term commitment to certified data repositories build the basis for the successful democratization of data. In the field of glacier data, this balancing act has so far been successfully achieved through joint collaboration between data repository institutions, data providers, and data users. However, we also note an unequal allotment of funds for data creation and projects using the data, and data curation. Considering the importance of glacier data to answering numerous key societal questions (from local and regional water availability to global sea-level rise), this imbalance needs to be adjusted. In order to guarantee the continuation and success of GTN-G in the future, regular evaluations are required and adaptation measures have to be implemented.

Keywords: glacier data, maturity matrix assessment, data repositories, Essential Climate Variable (ECV), Global Terrestrial Network for Glaciers (GTN-G)

BACKGROUND

The amount of glacier data has increased significantly over the last two decades. In the year 2000, data from around a hundred glaciers with direct mass-balance measurements and from around 1,000 glaciers with annual observations of terminus fluctuations were available. Today, satellite data in combination with the Randolph Glacier Inventory (RGI) that became available in 2012 (Pfeffer et al., 2014), enables the observation of all 215,000 glaciers worldwide. As a result, further observational parameters have been included in glacier repositories, such as glacier area and volume changes, flow velocities, ice-thickness distribution and snow-covered areas. This presents new challenges for the international data centers that provide access to glacier data, and fosters new expectations from users. In parallel, the storage (archiving), documentation (metadata), and access to the data and related products have become much more complex.

The Global Terrestrial Network for Glaciers (GTN-G) is the framework for the internationally coordinated monitoring of glaciers in support of the United Nations Framework Convention on Climate Change (UNFCCC). The network, authorized under the Global Climate Observing Systems (GCOS), is jointly run by the World Glacier Monitoring Service (WGMS), the U.S. National Snow and Ice Data Center (NSIDC), and the Global Land Ice Measurements from Space initiative (GLIMS). GTN-G represents the glacier monitoring community and provides an umbrella for existing and operational data services and related working groups such as of the International Association of Cryospheric Sciences (IACS). This setting is largely responsible for its success.

In addition to qualitative data, such as photographs and maps, the GTN-G provides standardized observations on changes in mass, volume, area and length of glaciers with time (glacier fluctuations), as well as statistical information on the spatial distribution of perennial surface ice (glacier inventories) (Figure 1). Such glacier fluctuation and inventory data are high-priority key variables in climate system monitoring; they form a basis for hydrological modeling with respect to possible effects of global warming, and provide fundamental information in glaciology, glacial geomorphology, and quaternary geology. The increased amount of glacier data from the last decade has enhanced the understanding of geophysical processes, improved glacier-related modeling, and resulted in higher-confidence statements in the last report of the International Panel on Climate Change (IPCC, 2021). Beyond this, the data are needed for the development of sustainable adaptation strategies and related decision-making processes in glacierized mountain regions (Nussbaumer et al., 2017; Gärtner-Roer et al., 2019). These urgent demands are accompanied by equally urgent challenges, such as the rapidly increasing number of glacier observations from space that need to be managed in a functioning database infrastructure.

GTN-G facilitates free access to data through different channels, depending on the level of interest and detail required, and addresses issues such as the standardization of measurement methods. Most important, it gathers high-level information about and access to all available glacier datasets on one platform (https://www.gtn-g.ch/data_catalog/). This ensures that

all data are equally available for any user: findable, accessible, interoperable, and reusable, following the FAIR principles (Wilkinson et al., 2016). While the public or mountain tourists might use the “wgms Glacier App” for a quick overview of available glacier data, scientists typically access data offered by the glacier services using the GTN-G data browser or directly from the catalog listing of data collections held by member repositories. Decision makers make use of edited products such as reports or specific country profiles. Finally, journalists often approach the GTN-G or constituent services directly and ask for support in filtering the main messages out of the full database and in showing different perspectives. Thus, the different repositories serve different user communities and purposes. However, whereas the FAIR principles emphasize the needs of data users, the right of the data providers to be acknowledged should not be neglected. Acknowledgment is accomplished through versioning of the datasets, e.g., *via* digital object identifiers (DOIs). When users cite datasets and include a DOI, the DOI provides traceability between data creation and use. During the whole process, the proper citation of data origin must be followed and ideally should be controlled by repositories, journals, and funding agencies. However, such control mechanisms have yet to be established by the international community.

Each GTN-G dataset nicely reflects the history of glacier monitoring, which began in 1894 with the internationally coordinated systematic observations on glacier variation (Figure 1). The history mostly followed the overall paradigms in science: after empirical and theoretical investigations, focus was given on simulations and, more recently, on “big data.” For the long-term monitoring of environmental variables, continuous and standardized measurements are of highest priority. The *in situ* measurements, where methodology has changed little over the last 125 years, are fundamental to this long-term monitoring. On the other hand, in order to capture uniform information on a large scale (glacier distribution, changes in ice thickness), remote sensing data are indispensable. The rise of “big data” in glaciology is a direct result of the rapid increase in remote sensing techniques and corresponding data, as well as free data-access policies (e.g., Landsat; Wulder et al., 2012) and the availability of “analysis ready data,” for example pre-orthorectified satellite scenes in GeoTIF format that can be easily processed and analyzed.

With the increase in volume, timeliness, and variety of data, as well as variety of data users, access becomes ever more challenging and requires improved interfaces (Pospiech and Felden, 2012). Citizens increasingly use data from different sources (maps, tides, etc.) and glaciers all around the world can now be explored and measured without much effort. This has implications for the management and handling of monitoring datasets that affect data providers as well as data users. Hence it is time to critically reflect on the democratization of glacier data. In the context of this study we understand the term “democratizing” as the free access to glacier data for everyone. As this is an active verb, it implies the transition of a former “closed” system to a more “open” system, even if access to most glacier data has been open already for many years (WGMS, 1998). For the future, it is the process of proper

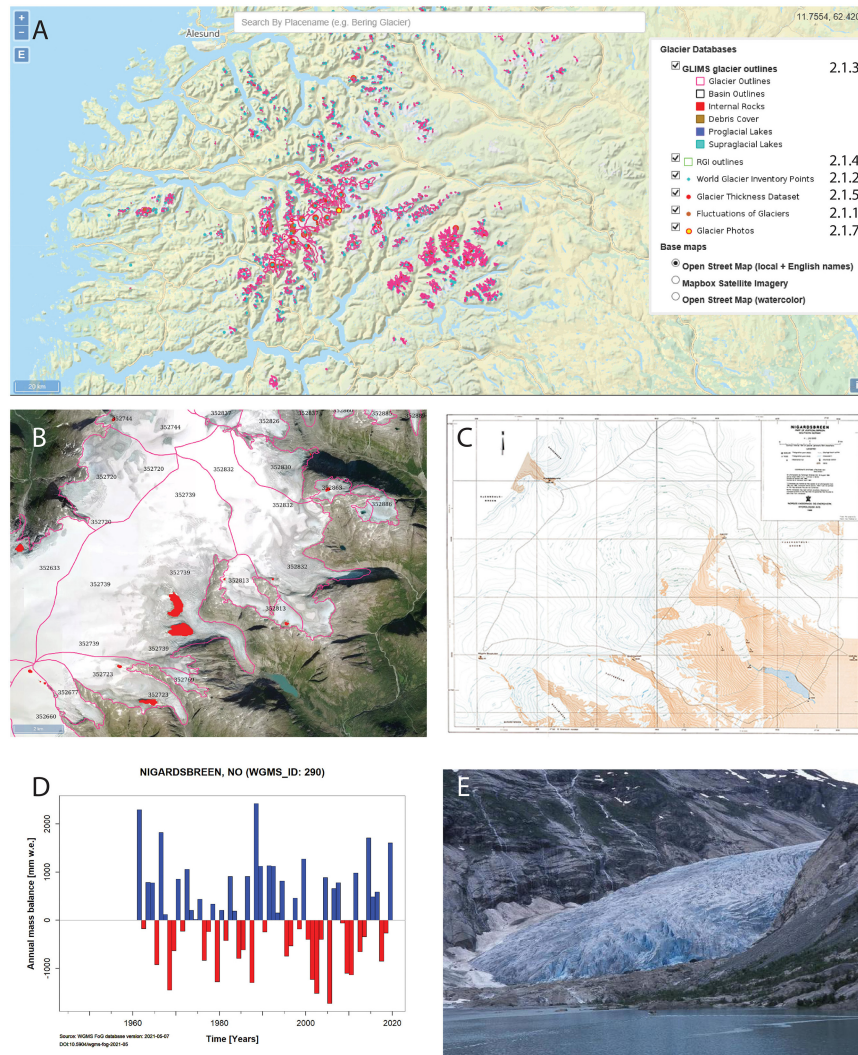


FIGURE 1 | (A) The GTN-G data browser (zoomed) showing available glacier data for a region in southern Norway. The legend shows the different datasets and the related sections in this paper where the datasets are described. Examples are given for Nigardsbreen, an outlet glacier of the Jostedalsgreen ice cap (Norway), as represented in different glacier datasets accessible via the GTN-G data browser: **(B)** GLIMS outlines as of 2006 (ID: 352739); **(C)** topographic map of Nigardsbreen as of 1998 (Norwegian Water Resources and Energy Directorate, NVE); **(D)** annual mass balance since 1960 (B. Kjøllmoen and colleagues, NVE; WGMS, 2021a); **(E)** photo of the glacier tongue as of August 3rd, 2000 (E. Roland; Glacier Photograph Collection).

data stewardship and international standardization that ensures the democratization of data. In this context, certification is provided by the CoreTrustSeal (<https://www.coretrustseal.org/>), an international, community based, non-governmental, and non-profit organization promoting sustainable and trustworthy data infrastructures.

We here systematically assess all seven GTN-G datasets with a focus on data preservability, accessibility, usability, production sustainability, quality assurance, quality control, quality assessment, transparency/traceability, and integrity, as described by Peng et al. (2015). The individual performance is analyzed with regard to the historical development and the current funding situation of individual datasets, but also with regard to each dataset's significance and function for environmental monitoring and related decision-making procedures. Particular challenges are stressed and suggestions

for solutions are provided by good-practice recommendations. During this process, the requirements of both data providers and data users are considered. Expectations from and ambitions of GTN-G are also formulated, as they direct the way toward the further democratization of glacier data.

DATA AND METHODS

Description of Datasets Available Within the GTN-G

Internationally coordinated collection and distribution of standardized information about glaciers was initiated in 1894 and is, since 1998, coordinated within GTN-G. Since 2008, an international steering committee coordinates, supports, and advises the operational bodies responsible for the international glacier monitoring, which are the WGMS, the NSIDC, and

GLIMS (Zemp, 2011). GTN-G ensures (1) the integration of the various operational databases and (2) the development of a one-stop web interface to these databases. The datasets all have different purposes, formats and histories, reflecting the history of glaciological science (Figure 2). By joint effort, consistency and interoperability of the different glacier databases has had to be developed; the different historical developments and methodological contexts of the datasets are challenges for linking individual glaciers throughout the databases.

For the analysis of the data, the interoperability with web-based services (e.g., cloud services) need to be improved. So far, most of the glacier datasets can be downloaded directly in their entirety and can be integrated into a programmatic local or cloud-based workflow. However, the linking between different GTN-G datasets is not very mature and urgently needs to be developed further. Developed in 2010 and updated since, a map-based web interface spatially links the available data and provides data users a fast overview of all available data (https://www.gtn-g.ch/data_browser/; see Figure 1). The interface was adapted for GTN-G from one developed for the constituent NASA-sponsored GLIMS initiative. It provides fast access to information on glacier outlines from about 215,000 glaciers mainly based on satellite images, length-change time series from 2,581 glaciers, glaciological mass-balance time series from 482 glaciers, geodetic mass-balance series from 37,446 glaciers, special events (e.g., hazards, surges, calving instabilities) from 2,747 glaciers, as well as more than 25,000 photographs (RGI Consortium, 2017; National Snow and Ice Data Center, 2021; WGMS, 2021b). By choosing the browser layer for a particular dataset one can quickly see the spatial distribution of that dataset. Whereas some datasets are fed continuously by an active community (such as the FoG (Fluctuations of Glaciers) and GLIMS datasets), others are created on an *ad-hoc* basis (GlaThiDa glacier thickness database and RGI dataset), have a random community (Glacier Photograph Collection, Glacier Map Collection) or have been discontinued (World Glacier Inventory, WGI).

The spatio-temporal coverage of the different datasets varies largely, because of their individual histories. For the *in situ* data there is a significant spatial bias toward the Northern Hemisphere, in particular to Europe and North America, whereas the Andes and Antarctica are underrepresented. In GlaThiDa, the largest spatial gaps persist in Asia, the Russian Arctic, and the Andes. With the recent developments in satellite remote sensing of the cryosphere, the extended sharing of data, and the free availability of a globally complete baseline glacier inventory (the RGI), near global coverage has been achieved for many datasets during the last decades (e.g., Farinotti et al., 2019; Hugonnet et al., 2021). Other temporal gaps in the datasets are related to the limited lifetime of individual projects or institutions. In addition, political crises can have a direct influence on the long-term continuation of data series. An assessment of national glacier distribution and changes, delineating also spatio-temporal gaps, is provided in Gärtner-Roer et al. (2019).

Fluctuations of Glaciers

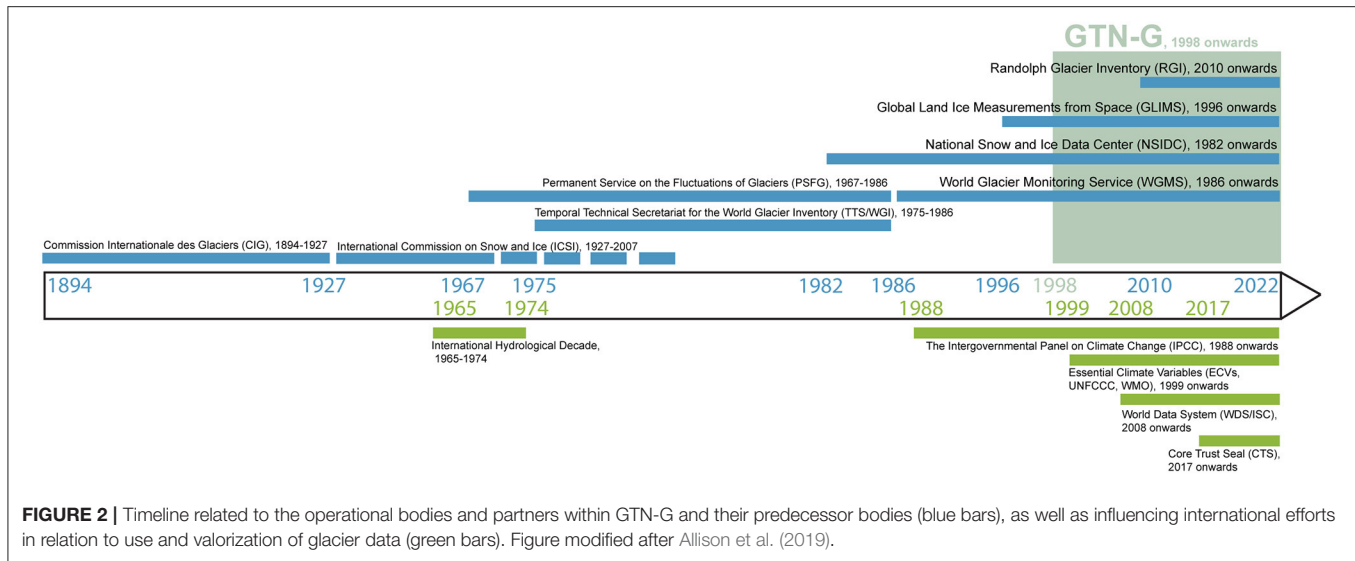
Internationally collected and standardized data on changes in glacier length, area, volume, and mass, based on *in situ*

and remotely sensed observations, as well as on model-based reconstructions, are compiled in the Fluctuations of Glaciers (FoG) database. The standardized compilation and free dissemination of glacier data from all over the world, as undertaken by the WGMS and its predecessor organizations, are a major contribution to international initiatives and bodies such as the United Nations Framework Convention on Climate Change (UNFCCC) and the Intergovernmental Panel on Climate Change (IPCC) (Figure 2). Since the beginning of coordinated glacier monitoring, the collected data have been published in written reports. The first reports were written in French, but from 1967 on, all reports are published in English (see https://wgms.ch/literature_published_by_wgms). The comprehensive FoG reports represented the backbone of the scientific data compilation, which comes with full documentation on principal investigators, national correspondents, their sponsoring agencies, and publications related to the reported data series. These reports, issued every 5 years, were complemented by the bi-annual Glacier Mass Balance Bulletin, which presented the data in summary form for non-specialists through the use of graphic presentations rather than as purely numerical data. In 2015, these two publication series were merged into the “Global Glacier Change Bulletin” series with the aim of providing an integrative assessment of worldwide and regional glacier changes at two-year intervals. Beyond these synthesis reports, the FoG data are accessed by downloadable files of past and current versions since 2008 (https://wgms.ch/data_databaseversions/), direct visualizations *via* the FoG Browser (<https://wgms.ch/fogbrowser/>), and the “wgms Glacier App” for mobile devices (<https://wgms.ch/glacierapp/>).

With the inclusion of near real-time measurements at high temporal resolution (e.g., hourly data) for selected study sites and the increasing amount of satellite-derived observations (number of records evolved from a few hundred to more than 200,000 glaciers), the database experienced growing pains. In order to address these challenges, plans for migration to advanced database structures are currently under development.

World Glacier Inventory

The WGI was planned as a snapshot of glacier occurrence on Earth during the second half of the 20th century. In 1976, the United Nations Environment Programme (UNEP), through its Global Environment Monitoring System (GEMS) started supporting activities of a Temporary Technical Secretariat for the World Glacier Inventory (TTS/WGI) established at the Geography Department of ETH (Eidgenössische Technische Hochschule) Zurich. Detailed and preliminary regional inventories were compiled all over the world. From these inventories, statistical measures of the geography of glaciers could be extracted. The WGI completed and updated earlier compilations (e.g., by Mercer, 1967 and Field, 1975). Instructions and guidelines for the compilation of standardized glacier inventory data were developed by UNESCO/IASH (1970), Müller et al. (1977), Müller (1978), and Scherler (1983). The publication of the WGI report (WGMS, 1989) presented the status at the end of 1988, and is the first such compilation to give a systematic global overview. It contains information for



approximately 130,000 glaciers. Inventory parameters include geographic location, area, length, orientation, elevation, and classification. The WGI is based primarily on aerial photographs and topographic maps with most glaciers having one data entry only. Hence, the dataset can be viewed as a snapshot of the glacier distribution in the second half of the 20th century. An update of the WGI was performed in 2012 (WGMS, and National Snow and Ice Data Center, 2012).

The data collection presents a fairly complete, albeit preliminary, picture of the world's glacierized regions at the given time. The WGI database is stored both at WGMS in Zurich and in the National Snow and Ice Data Center's NOAA collection, part of the World Data Center for Glaciology in Boulder, Colorado. It is most easily accessed through the website of GTN-G (www.gtn-g.org). The WGI database is searchable by glacier ID, glacier name, or latitude/longitude (as well as other parameters) using the main "Search Inventory" interface. In addition, the "Extract Selected Regions" interface can be used.

It was the sincere wish of organizations and people who have been involved in WGI activities over the years that the information in the publication, together with the data available in the database, be of service to scientists and decision makers concerned with various applications of glacier data both then and in the future (WGMS, 1989). For instance, it was suggested that the information available within the WGI together with other data provided by the WGMS could be usefully applied in studies of the impact of a global warming on the availability of water resources in frozen form, particularly in semi-arid and arid regions bordering glacierized areas. Inventory data had already proven useful for estimating precipitation amounts in some mountainous regions where stations for direct measurements are difficult to establish, and it was expected they would be used for the same purpose in many more regions (WGMS, 1989).

Independent of the high scientific value of the glacier information stored in the WGI, it has some disadvantages when considering today's applications. The foremost problem is its

storage as point information. The shapes and the extents of the glaciers to which the data belong are unknown. It cannot be used for change assessment or any application that requires glacier outlines. The technological revolution in the 1990's providing Geographic Information Systems (GIS), digital elevation models (DEMs) and satellite data covering nearly each region in the world with glaciers, has made it possible to generate, store and manipulate related vector data. As a consequence, the GLIMS database (see Global Land Ice Measurements From Space) was initiated at the turn of the century, superseding the WGI. The compilation of a near-globally complete dataset of glacier outlines as available from the RGI (see Randolph Glacier Inventory) was, however, only possible once free access to orthorectified satellite data, DEMs, and GIS environments was in place.

Global Land Ice Measurements From Space

The GLIMS glacier database (GLIMS Consortium, 2005) contains multi-temporal outlines of glaciers in a vector format with additional data about each glacier (e.g., name, area, length or mean elevation). All data are stored in a PostGIS relational database, providing support for geographic objects allowing location queries. It emerged from the increasing need for improved calculation of glacier changes and glacier-specific assessments, which were impossible using the point data provided by the WGI (see above). As the WGI and its extended format WGI-XF (Cogley, 2009) was still spatially incomplete, there was also an urgent need to obtain complete global coverage, i.e., to have outlines from all glaciers in the world rather than just 2/3. At the inception of GLIMS in 2010 it was still not known how many glaciers we had on Earth, where they were located and how large they were. Accordingly, all calculations concerned with regional scale hydrology in mountain regions or global scale sea-level rise were highly error prone. With the free availability of multispectral images at 15 m spatial resolution from the ASTER sensor onboard the Terra satellite (after its launch in 1999), the dream of a global glacier database suddenly became realistic. Data

acquisition requests for ASTER were prepared (Raup et al., 2000) and a geospatial database was created (Raup et al., 2007). The relational database included everything that possibly could be derived from a satellite image and slowly filled over the years.

Whereas algorithms for automated mapping of clean glacier ice were already established at that time (e.g., Bayr et al., 1994; Paul et al., 2002), two major bottlenecks hindered rapid and efficient data processing: (a) debris-covered glacier parts were not included and had to be delineated manually and (b) image analysts had to manually orthorectify all ASTER scenes. With no money available for such time-consuming activities both could only be performed as a part of funded research projects that mostly analyzed small regions (e.g., Paul and Kääb, 2005). Fortunately, the opening of the Landsat archive in 2008 (Wulder et al., 2012) suddenly provided free access to all Landsat scenes in an already orthorectified format and obviated the need to use manually orthorectified ASTER imagery. This encouraged glacier mapping over larger regions (Bolch et al., 2010; Frey et al., 2012; Rastner et al., 2012; Guo et al., 2015) which filled the GLIMS database with more, better quality and also multi-temporal data. At the time of writing, the GLIMS database hosts approximately 300,000 glacier outlines (including perennial snow patches), i.e., 40% of the 215,000 glaciers have multi-temporal outlines. The data have been widely used for a range of hydrological and glaciological applications. The datasets stored in GLIMS also formed the base for the compilation of a first globally complete single snapshot inventory (RGI; see next section).

Randolph Glacier Inventory

The RGI was born from two ideas: to have (i) an easily accessible temporal snapshot of glacier extents available that is (ii) globally complete, i.e., there is one outline for each glacier in the world with the relevant attribute information. This idea was motivated primarily by the preparation of IPCC AR5, where a clear need for such a dataset was communicated to the glaciological community to improve the assessment of glacier-related questions (e.g., their contribution to sea-level rise) compared to IPCC AR4. With glacier outlines from the GLIMS database and a special community effort in glacier mapping (for details see Pfeffer et al., 2014), first versions of this dataset were created and provided for the global-scale calculations presented in IPCC AR5 (Vaughan et al., 2013). Given the limited time available for finalizing the product, shortcomings in quality were accepted, noting that the outlines were produced for global to continental scale assessments rather than regional or local ones. Over time, the RGI was continuously improved (version 6.0 appeared in 2017) and the regionally most complete datasets were collected and combined for the best possible product.

Whereas, the initial effort to get all data together in a consistent format was enabled by a couple of engaged individuals, the current effort for compilation of a further improved RGI (version 7) is coordinated by a dedicated IACS working group (<https://cryosphericciences.org/activities/working-groups/rgi-working-group/>) that is organizing and structuring the related work. A detailed technical specification about RGI contents, its development over time, and all its contributors is available in the form of a Technical Note

from the RGI web page (https://www.glims.org/RGI/00_rgi60_TechnicalNote.pdf). The RGI is split into 19 first order regions, each having its own glacier outlines shapefile and hypsometric data file. When summed up, it contains about 215,000 glacier entities covering an area of more than 723,000 km² (excluding glaciers on the Antarctic Peninsula).

The RGI has likely become one of the single most important datasets for glaciological and hydrological research. It is widely accepted as the best possible dataset for large scale applications and the number of studies using it might exceed 1000. The related study by Pfeffer et al. (2014) describing version 3.2 in detail is now the most cited publication in the *Journal of Glaciology*. The intense use of the dataset is also a main reason for ongoing efforts to further improve it, being careful not to lessen its usefulness. For the new version 7 of the RGI it was decided to bring the individual datasets closer to the year 2000 (e.g., to facilitate mass-balance calculations starting with the SRTM or ASTER-derived DEMs from 2000) and swap out datasets with known problems (e.g., too much seasonal snow mapped as glaciers in the Andes) for “better” ones.

A dataset such as the RGI is never perfect nor complete. Whereas, obvious errors such as too much seasonal snow being mapped as glaciers, wrongly mapped debris-covered glaciers or (frozen) lakes or missing glacier parts due to clouds can be detected and corrected, variability in the interpretation (is this a rock glacier or a debris-covered glacier?, where is the drainage divide?) or topological issues (is this one ice cap or many glaciers?) are much harder to address. They will persist in future versions of the RGI as there is no unique right or wrong answer to these questions. In the end, a user of the dataset can always consult the larger GLIMS database when searching for an alternative interpretation of glacier extent or the timing of the outline does not fit to the intended application. Apart from the glacier mapping itself that should become more precise over time as increasingly high resolution satellite images (e.g., Sentinel-2) are available (Paul et al., 2016), the extraction of a “new” RGI version from the GLIMS database is not a button-press application but requires considerable effort. It is yet unclear if funding will be available for this in the future. The creation of RGI version 7 is largely automated now so that future RGI versions can be extracted from the GLIMS glacier database according to a set of pre-scribed criteria with limited effort. However, due to topological inconsistencies and the different internal handling of glacier datasets the creation of this automation has been time-consuming.

Glacier Thickness Database

GlaThiDa is the only worldwide database of glacier ice thickness observations, and thus plays an important role in studies of glacier ice volumes and their potential sea-level rise contributions (e.g., Farinotti et al., 2017; MacGregor et al., 2021). The measurements are compiled from literature reviews (e.g., Gärtner-Roer et al., 2014), imported from published datasets, or submitted by researchers in response to calls for data. While major versions of GlaThiDa are archived at the WGMS (e.g., <https://doi.org/10.5904/wgms-glathida-2020-10>), the dataset is developed online as a version-controlled “git”

repository (<https://gitlab.com/wgms/glathida>). The development environment (described in Welty et al., 2020) automatically records changes to the dataset, continuously checks the integrity of the dataset, and provides an interface for bug reports, feature requests, and other community dialogue. Although a few suspicious ice thickness measurements have been flagged manually, source data are not automatically checked for plausibility, and in some cases they may be very wrong (e.g., <https://gitlab.com/wgms/glathida/-/issues/25>). Additional checks could be developed to automatically flag data that are inconsistent with neighboring measurements, modeled ice thicknesses (e.g., Farinotti et al., 2019), or glacier outlines (e.g., GLIMS, RGI).

Glacier Map Collection

Many glacier maps were published in the FoG reports between 1967 and 2012. They often show individual glaciers and their spatio-temporal changes in very detailed mappings, some of them with outstanding quality. Several glaciers, e.g., Lewis glacier (Kenya), were mapped repeatedly over many decades. This additional dataset complements the FoG database with more qualitative and comprehensive environmental information. To enable a direct access and use, the maps were digitized and made available online in 2018 (wgms.ch/products_fog_maps). Sporadically, additional maps (newly created and digitized old maps) are added to the collection.

Glacier Photograph Collection

The Glacier Photograph Collection (National Snow and Ice Data Center, 2021) is an online (https://nsidc.org/data/glacier_photo/search/), searchable database of digital photographs of glaciers from around the world, some dating back to the mid-19th century, which provide a historical reference for glacier extent. The photos are either scanned from physical objects such as photographic prints or slides or they originated in digital form from a digital camera. As of May 2022, the database contains over 25,500 photographs. Most of the photographs are of glaciers in the Rocky Mountains of North America, the Pacific Northwest, Alaska, and Greenland. However, the collection does include a smaller number of photos of glaciers in Europe, South America, the Himalayas, and Antarctica. The collection includes a number of sub-collections or Special Collections that are distinguished in some way. For example, there is a special collection of Repeat Photography of glaciers that provides a unique look at changes in glaciers over time. These photographs constitute an important historical record, as well as a data collection of interest to those studying the response of glaciers to climate change. Educators use the photographs frequently and artists have found inspiration in the photographs.

The collection is maintained by NSIDC. New photographs are submitted from a wide community and are added to the database sporadically. The collection is accessible on the NSIDC website, using a detailed search interface that allows request for regional or national data and individual glaciers, as well for specific years and single photographs.

Data Stewardship Assessment: Parameters

The assessment of the different GTN-G datasets is performed by the compilation of individual maturity matrices. Each matrix compiles all information on preservability, accessibility, sustainability, quality, reproducibility, and integrity of the data and metadata in each dataset following the approach by Peng et al. (2015, 2019), which is explained in more detail below. The assessment is based on the conceptual model and the related scoreboard presented by Peng et al. (2015). The individual evaluation criteria are slightly adapted to the “language” of glaciologists (see **Table 1**). The individual assessments of the seven datasets are compiled in a separate score table (**Table 2**).

This maturity scale contains nine key components. For each dataset, a maturity score from 1 to 5 is assigned, representing five levels of maturity. The levels range from Level 1, corresponding to a dataset that was developed *ad-hoc* and that is not managed, to Level 5, representing a dataset that is optimally managed and developed on the long-term and that is externally audited (**Table 1**). The assessment was compiled by the authors of the present paper who are managers of the different GTN-G datasets. This expert evaluation is characterized by a multi-step approach. First, each manager completed a full assessment of their respective dataset (self-assessment) based on their interpretation of the criteria as compiled in **Table 1** and their reading of the original work by Peng et al. (2015). Second, as each dataset is run by several experts, the individual assessments were discussed with the other people responsible for the dataset in an iterative process to achieve a consensus. In a last iteration, this consensus assessment was presented to and discussed with the managers of the other datasets, representing also the GTN-G Executive Board. Related key words for the assignment of the score are given in **Table 1** for each assessment criteria (key component). In the following, the assessment criteria are described from a more glaciological perspective:

Preservability

Are there any archiving standards (e.g., CoreTrustSeal) for the dataset? Is there redundancy? Do the archiving processes follow certain standards? Is there any predictive planning for future changes?

Accessibility

Are the data publicly available? Do the data services follow the sense of community standards? Is there additional dissemination of data products to enhance data accessibility for different user groups?

Usability

Data format: is it standard/non-standard? Are there interoperable formats? Is the available metadata adequate and in a usable form? Are the data and metadata sufficiently documented? Is there any need for specific knowledge to use the data? Are there online capabilities available, such as visualizations or a product user guide?

TABLE 1 | Maturity scale applied for the assessment of the GTN-G datasets [modified after Peng et al. (2015, 2019)].

| Maturity scale / GTN-G data sets (across) | Level 1 | Level 2 | Level 3 | Level 4 | Level 5 |
|--|--|--|---|--|--|
| Key components (below) | Ad-hoc not managed | Minimal managed limited | Intermediate managed defined partially implemented | Advanced managed well-defined fully implemented | Optimal level 4 + measured controlled, audit |
| Preservability: (The state of being preservable) | Any data storage location data only (no metadata) | Non-designated repository data storage redundant limited archiving metadata | Designated repository (e.g., CoreTrustSeal) data storage redundant community-standard archiving metadata limited archiving standards apply | Designated repository (e.g., CoreTrustSeal) data storage redundant community-standard archiving metadata community archiving standards apply | Designated repository (e.g., CoreTrustSeal) data storage redundant community-standard archiving metadata community archiving standards apply archiving process monitored and audited planned future data archiving |
| Accessibility: (The state of being searchable and accessible publicly) | Data not publicly available data access person-to-person | Data is publicly available direct file download possible data searchable online (on dataset level) | Data is publicly available direct file download possible non-standard data service provided limited data server performance data searchable online (on granule/file level) limited search metrics | Data is publicly available direct file download possible community-standard data service provided enhanced data server performance data searchable online (on granule/file level) community search metrics internal dissemination report | Data is publicly available direct file download possible community-standard data service provided enhanced data server performance data searchable online (on granule/file level) community search metrics dissemination report available online planned future data accessibility |
| Usability: (The state of being easy to use) | Specific knowledge required no documentation online | Non-standard data format limited documentation online | Community-standard data format (incl. Metadata) documentation online | Community-standard data format (incl. Metadata) documentation online basic data characterization online | Community-standard data format (incl. Metadata) documentation online enhanced data characterization online (e.g., visualization) community metrics of data characterization online external ranking |
| Production sustainability: (The state of data production being sustainable and extendable) | <i>Ad-hoc</i> initiative no deliverables existing | Short-term initiative individual commitment by PI's | Medium-term initiative institutional commitment | Long-term initiative (program) institutional commitment product improvement process in place | Long-term initiative (program) national or international commitment planned product improvement |
| Data quality assurance (DQA): (The state of data quality being assured) | DQA procedure unknown or inexistent | DQA procedure random DQA procedure not defined and documented | DQA procedure defined, documented and partially implemented | DQA procedure well-documented, fully implemented and available online limited DQA metadata available | DQA procedure well-documented, fully implemented, available online, monitored and reported community-standard DQA metadata available external review of DQA |
| Data quality control/monitoring: (The state of data quality being controlled and monitored) | No quality monitoring of data and metadata | Limited monitoring of data and metadata | Regular monitoring of data and metadata, not automatic | Fully automatic monitoring of data and metadata following community standards consistency checks | Fully automatic monitoring of data and metadata following community standards consistency checks provider/user feedback in place planned future data quality control |
| Data quality assessment: (The state of data quality being assessed) | Method and theoretical basis assessed | Method and theoretical basis assessed research product assessed | Method and theoretical basis assessed research product assessed operational product assessed | Method and theoretical basis assessed research product assessed operational product assessed quality metadata assessed | Method and theoretical basis assessed research product assessed operational product assessed quality metadata assessed assessments performed on recurring basis and following community standards external evaluation |

(Continued)

TABLE 1 | Continued

| Maturity scale / GTN-G data sets (across) | Level 1 | Level 2 | Level 3 | Level 4 | Level 5 |
|--|---|---|--|---|--|
| Key components (below) | Ad-hoc not managed | Minimal managed limited | Intermediate managed defined partially implemented | Advanced managed well-defined fully implemented | Optimal level 4 + measured controlled, audit |
| Transparency/traceability: (The state of being transparent, trackable, and traceable) | Limited product information available knowledge transfer person-to-person | Product information available in the literature | Product information available | Product information online Unique Object Identifier (OID) assigned Digital Object Identifier (DOI) assigned | Complete data provenance online Unique Object Identifier (OID) assigned Digital Object Identifier (DOI) assigned planned future data traceability |
| Data integrity: (The state of data integrity being verifiable) | Unknown or no data ingest integrity check | Data ingest integrity verifiable | Data ingest integrity verifiable data archive integrity verifiable | Data ingest integrity verifiable data archive integrity verifiable data access integrity verifiable | Data ingest integrity verifiable data access integrity verifiable data authenticity verifiable (hash codes, digital signatures) performance of data integrity checks |

Production Sustainability

To what extent is there a commitment and stewardship, from individuals (e.g., Principal Investigators) to organizations/services (e.g., WGMS)? What is the rating of the dataset, ranging from *ad-hoc* initiatives (with or without deliverables) to long-term programs secured through national or international funding?

Data Quality Assurance

Is a DQA procedure implemented? Is the procedure manual or automated? Is there sufficient documentation of the DQA? Are there any reports about the DQA, according to community standards and with external review?

Data Quality Control/Monitoring

Is the data quality controlled and monitored based on a regular sampling and analysis? Is there a systematic and/or an automatic procedure? Are there regular consistency checks following community standards? Are provider and user feedback mechanisms in place?

Data Quality Assessment

Are there quality reports for methods and results? Is there sufficient metadata about quality assessment? Is there an assessment on a recurring basis? Is there an external evaluation?

Transparency/Traceability

Is there (online) product information available? Is the data provenance sufficiently documented and are there related operational algorithms? Are the data governance mechanisms online available? Is all information important for reproducibility available?

Data Integrity

Are there integrity checks? How do they perform, are they verifiable? Integrity checks should address the following: ingestion of data, data archiving, data access, data authenticity. Is there a monitoring and reporting of the performance of data integrity checks?

RESULTS: PERFORMANCE OF THE GTN-G DATASETS

Table 2 summarizes the performance of the evaluated GTN-G datasets. A score from 1 to 5 was assigned for each key component, which is explained with the comments given in Table 1.

Fluctuations of Glaciers

The FoG database performs between advanced and optimal. It is a designated repository for glacier fluctuations data following standards of the glaciological community and key standards regarding archiving quality and security. Data are accessible through different channels. Data quality assurance (DQA) is manually ensured, but not automatically enforced. Data integrity checks are not fully automatic. Each version of the dataset is identified by its own DOI and the provenance of the data is documented in detail both in the metadata and the database itself.

TABLE 2 | Summary matrix with the performance of the seven datasets available within GTN-G.

| Maturity scale / GTN-G data sets (across) | FoG | WGI | GLIMS | RGI | GlaThiDa | GMC | GPC |
|--|-----|-----|-------|-----|----------|-----|-----|
| Preservability: (The state of being preservable) | 4 | 4 | 3 | 3 | 4 | 2 | 4 |
| Accessibility: (The state of being searchable and accessible publicly) | 4 | 3 | 4 | 4 | 3 | 2 | 3 |
| Usability: (The state of being easy to use) | 5 | 4 | 4 | 4 | 4 | 2 | 3 |
| Production sustainability: (The state of data production being sustainable and extendable) | 5 | 3 | 3 | 3 | 3 | 1 | 4 |
| Data quality assurance (DQA): (The state of data quality being assured) | 3 | 3 | 2 | 2 | 3 | 1 | 2 |
| Data quality control/monitoring: (The state of data quality being controlled and monitored) | 3 | 2 | 3 | 3 | 4 | 2 | 2 |
| Data quality assessment: (The state of data quality being assessed) | 5 | 2 | 2 | 2 | 2 | 1 | 2 |
| Transparency/traceability: (The state of being transparent, trackable, and traceable) | 5 | 4 | 2 | 2 | 4 | 3 | 3 |
| Data integrity: (The state of data integrity being verifiable) | 3 | 3 | 3 | 3 | 2 | 1 | 3 |

A score from 1 to 5 is given for each key component. The traffic-light colors give an additional hint to the maturity of the single datasets: green colors: good performance, yellow: medium performance, orange: limited performance.

TABLE 3 | List of all GTN-G datasets with their URL (Uniform Resource Locator) and citation.

| Dataset | URL | Citation |
|--|---|---|
| Fluctuations of Glaciers (FoG) | https://dx.doi.org/10.5904/wgms-fog-2021-05 | WGMS, 2021a: Fluctuations of Glaciers Database. World Glacier Monitoring Service, Zurich, Switzerland. doi: 10.5904/wgms-fog-2021-05 |
| World Glacier Inventory (WGI) | https://nsidc.org/data/glacier_inventory/index.html | WGMS, and National Snow and Ice Data Center (2012) World Glacier Inventory. Compiled and made available by the World Glacier Monitoring Service, Zurich, Switzerland, and the National Snow and Ice Data Center, Boulder CO, USA. Digital Media |
| Global Land Ice Measurements from Space (GLIMS) Initiative | https://www.glims.org | GLIMS and National Snow and Ice Data Center (2021): Global Land Ice Measurements from Space glacier database. Compiled and made available by the international GLIMS community and the National Snow and Ice Data Center, Boulder CO, U.S.A. doi: 10.7265/N5V98602 |
| Randolph Glacier Inventory (RGI) | https://www.glims.org/RGI/index.html | RGI Consortium (2017): Randolph Glacier Inventory – A Dataset of Global Glacier Outlines: Version 6.0: Technical Report, Global Land Ice Measurements from Space, Colorado, USA. Digital Media. https://doi.org/10.7265/N5-RGI-60 |
| Glacier Thickness Database (GlaThiDa) | https://www.gtn-g.ch/data_catalog_glathida/ | GlaThiDa Consortium (2020): Glacier Thickness Database 3.1.0. World Glacier Monitoring Service, Zurich, Switzerland. doi: 10.5904/wgms-glathida-2020-10 |
| Glacier Map Collection (GMC) | https://wgms.ch/products_fog_maps/ | WGMS (2018): Glacier Map Collection (GMC), World Glacier Monitoring Service, Zurich, Switzerland. doi: 10.5904/wgms-maps-2018-02 |
| Glacier Photograph Collection (GPC) | https://nsidc.org/data/glacier_photo/ | National Snow and Ice Data Center (2021): Glacier Photograph Collection, Version 1. Boulder, Colorado USA. NSIDC: National Snow and Ice Data Center. https://doi.org/10.7265/N5/NSIDC-GPC-2009-12 |

For this advanced performance, the FoG database was already certified as trustworthy repository by CoreTrustSeal in 2019.

World Glacier Inventory

The WGI dataset performs between intermediate and advanced. It is a well-managed dataset with clearly defined aims and purposes. Lower scores stem from the data quality control/monitoring (limited monitoring of data and metadata) as well as the data quality assessment (assessment is performed

of the research product, but not of the dataset itself). As this dataset represents a snapshot from 1989, with an update from 2012, data curation is currently non-existent. Hence, the overall performance is low.

Global Land Ice Measurements From Space

The GLIMS database performs with an overall score of intermediate. While the database is accessible through an

enhanced data server with provided dissemination metrics according to community standards, a clear backdrop is the largely voluntary basis of data provision by scientists from all over the world. This leads to data contributions happening by chance from research projects, with a wide range of interpretation of glacier extent and limited control on data quality and uncertainty assessment. Although several guidelines exist for the community (Raup and Khalsa, 2007; Paul et al., 2009), various quality checks have to be performed before data ingest. This includes file formats, completeness of metadata, topologic errors, location errors, outline quality, etc. Despite automated tools being available for parts of this work, this still requires effort, especially for less standard data formatting.

Randolph Glacier Inventory

The dataset performs similar to the GLIMS database, with an overall score of intermediate. Data quality assurance (DQA) takes place on an *ad-hoc* basis, is not systematic, and is dependent on the data provider. Standard checks for data quality control are implemented but not documented. Data transparency is low, because product info can only be found in the literature that documents the submitted data (but it is available and citable).

Glacier Thickness Database

The dataset achieves an overall performance between intermediate and advanced. There is a designated repository for the dataset (WGMS), which stores all versions and their metadata, and ensures public and direct access to the files. Data and metadata are machine-readable following community standards. Data quality checks are performed systematically and automatically, and are all documented either in the metadata or in the source code. A lower performance stems from relatively low production sustainability, as there is only a medium-term commitment from the data repository to further develop the dataset. Only cursory data integrity checks are performed, but all changes to the files are tracked in a version-controlled (git) repository.

Glacier Map Collection

This collection performs between minimal and *ad-hoc* for management of the dataset, which currently consists of *ad-hoc* initiatives (though regular inclusion of new maps in former times). Data are accessible online, but there is only limited documentation of the data itself. DQA procedures are random and only the method and its theoretical basis are assessed. There are no data integrity checks performed.

Glacier Photograph Collection

The collection is preserved at a designated repository (NSIDC) with well-formed dataset and file-level metadata, following high community-archiving standards. There is a direct and public access of the data, with some search metrics provided. Long-term commitment by the data repository ensures the production sustainability. DQA procedures are performed but not defined or documented. Data product information is available in a user guide and data integrity checks are in place. This leads to an overall performance of intermediate (advanced in a few criteria).

DISCUSSION OF OPPORTUNITIES AND CHALLENGES

The maturity matrix approach (Peng et al., 2015) applied in this study allows a clear and comprehensive assessment of the individual glacier datasets, as well as a cross-comparison to other datasets. Similar assessment schemes for maturity matrices are available (Bates and Privette, 2012; Albani and Maggio, 2020; CEOS, 2020), often with very similar parameters as they are predominantly applied in environmental sciences. For example, the European Organization for the Exploitation of Meteorological Satellites (EUMETSAT) uses the maturity matrix to assess the maturity of climate data records and the development of Essential Climate Variables (ECVs). EUMETSAT applies the systematic approach by Bates and Privette (2012) to assess if the data record generation follows best practices in the areas of science, information preservation, and usage of the data. This approach was also used when preparing the Copernicus Climate Change Service (C3S) and assessing the needs for full access to standardized climate change data. In this C3S context, the FoG and RGI glacier datasets were also evaluated regarding the availability and quality of metadata, user documentation, uncertainty characterization, public access, and usage. A comparison of the C3S assessment with the outcome of this study reveals a congruent performance.

The assessment of glacier datasets showed that most of the datasets perform on an intermediate level. Given the individual significance of the datasets, the most important ones, when it comes to basic data on glacier distribution and glacier changes, are managed on a long-term perspective, but have only limited funding.

Historical Development

The current state of the GTN-G datasets can largely be explained through their historical development. The glacier fluctuation dataset (FoG) traces back to the end of the 19th century, when the worldwide coordination of glacier monitoring was initiated. With time, the uninterrupted continuation of the data collection has become a strong argument to further institutionalize the collection of glacier data. This led to the formation of the Permanent Service on the Fluctuation of Glaciers (PSFG) in 1967, under the umbrella of international auspice organizations, and later in 1986 to the formation of the WGMS. The commitment of the coordinators of this network as well as the dedication of many investigators and collaborators in turn helped to emphasize the achievements and positive reception of the services. Different challenges that emerged during that time had to be tackled, and different needs from data users, data producers, and international organizations have to be satisfied by the international data centers. As a consequence, this development is also reflected in the GTN-datasets as presented of today.

This can be seen in several examples. First, FoG emerged from simple length change measurements and later on included *in situ*, geodetic and point mass balances. Hence, the dataset has become more comprehensive, but this also needed more effort for maintenance and continued support. Second, GLIMS developed

from glacier outlines for individual regions to a globally near-complete and partly multi-temporal glacier inventory, forming the base for the RGI that reached completeness with a different data model independently and was in turn ingested into GLIMS to get it spatially complete. To maintain the RGI for the long-term, it is now also provided via GLIMS and the NSIDC. These examples show the independent development of each dataset while maintaining links to the other GTN-G datasets, such as GPC or GlaThiDa, which are steadily increasing in data richness.

We note that GTN-G has developed in a research environment and, hence, has never reached the support levels of operational monitoring networks such as within the WMO with its national meteorological services. In summary, there is a long history in glacier monitoring and so far, a good job was done, but with the increasing amount of data, the challenges and requirements (from different users) have changed and need to be tackled. This is only possible with data curation and stewardship.

Data Curation and Stewardship

The increasing amount of data from different sources pushes most storage systems to their capacity limits and require regular expansion of the hardware, including mirror sites. To ensure fast and long-term data access, constant updates of the software are required as well. Further, the increasing demand for direct access to most up-to-date information is a common desire of many data users. Therefore, data feeds, checks and updates must be carried out continuously. In parallel, proper dataset versioning is needed to guarantee traceability. Most of these challenges come with technical demands of increasing complexity.

While the NSIDC offers a rather large infrastructure for data repositories but has limited funding for active (hands-on) data curation, the WGMS is a small service that has its strengths in data analysis with a strong focus on one specific database but limited capacities to host additional datasets. Hence, for the different operational bodies, individual data curation strategies need to be set up, evaluated and revised on a regular basis and the responsible database manager(s) need to run consistent procedures of data archiving, access and quality checks. Regular training and exchange with colleagues from the glacier community would also be an advantage to take up current challenges quickly and become responsible data stewards. Following these procedures will professionalize the repositories, strengthen the data services, and serve the community of data providers and users optimally.

The best-practice measures mentioned above of course come at a price. In addition to upkeep of technical equipment (hardware, software), science officers and database managers need to be trained technically and substantively to ensure a qualified data processing chain. To bolster support for adequate technical equipment and staff training for the different data services, support is needed from higher-level agencies or international organizations. They are the only ones that can commit and contribute to the data services for the long run. Therefore, lobbyists are needed who communicate the recent shortcomings and challenges to the responsible decision makers. In the case of GTN-G, this task could be taken on by the Advisory

Board, as they know the glacier community sufficiently well and have the right contacts to international organizations.

Funding Situation

From the assessment, we noted a direct relationship between the scores of the datasets and the respective funding available to maintain it. Funding often comes from research projects that cover at most the next few years. In these cases, a long-term perspective is lacking, since follow-up projects that would ensure a direct continuation are often not guaranteed, or even dismissed due to the “lack of innovation.” Hence, existing structures first need to be sustained for a more long-term operation.

Within GTN-G, the funding situation currently looks as follows: the only dataset with dedicated long-term funding for data management is the FoG dataset (with 3 FTE (full time equivalent), Swiss GCOS 2021-24; C3S 312b 2020-21). In addition, the RGI runs on short-term funding (1 FTE, C3S 312b 2020-21), as does GLIMS (0.5 FTE, NASA Distributed Active Archive Center funding). The other datasets are updated on a more voluntary or *ad-hoc* basis without dedicated funding, although the WGI and the GPC are minimally maintained with support from the NOAA Cooperative Agreement with CIRES, NA17OAR4320101. In the future, *ad-hoc* data compilations, such as GlaThiDa, will be easier to fund, as they can build on existing structures and can be linked to scientific projects or sponsored by societies such as IACS. On the other hand, long-term monitoring necessitates a long-term commitment, which is more difficult to secure funding for. Running trustworthy repositories needs long-term security and perspective. Dedicated support and long-term commitment for certified data repositories build the basis for the successful democratization of data.

In the field of glacier data, this balancing act has so far been successfully achieved through joint collaboration between data repository institutions, data providers, and data users. However, the money spent on the data provider and user side for creating and working with the datasets (generally, scientific projects) is several orders of magnitude larger than the funds available for data curation. Hence, international organizations as well as national authorities must offer support and take responsibility on both sides. Most challenges can only be overcome in a financially safe and secure setting for data services and with the help of international standardization, as, for example, provided by the CoreTrustSeal.

For the GLIMS glacier database, the funding situation is too low to elevate its maturity score. GLIMS was started and maintained for some years on short-term (3–5 year) project funds, but has recently been folded into the NSIDC DAAC, funded by NASA. Current GLIMS activities are being performed mainly by one person at a 40% engagement, with other software developers contributing on occasion. The database has some issues that need to be improved to reach a higher standard, but without sufficient and sustained funding of the required experts this is difficult, or too slow. Given the importance of this database for many other multi-million-dollar projects, the limited funding available for keeping the database healthy and growing is more than shameful. We acknowledge, however, that this is also a result of the historical development up to the current explosion

of available datasets and recently changed user demands and possibilities (e.g., cloud computing).

Future Ambitions

In respectful view of the historical developments and the awareness of the recent challenges, the individual GTN-G data services need to take urgent actions. Minimum actions are required to simply keep the state-of-the-art. Far-reaching measures must be taken to secure the future of data services and their benefits for the entire community and to serve different stakeholders—experts, policy makers, the interested public and journalists.

As mentioned in Data Curation and Stewardship, individual data curation strategies are needed for the operational bodies of GTN-G. Trained database managers will have to organize and monitor the implementation of this strategy, and run consistent procedures for data archiving and to perform access and quality checks on a regular basis. In addition, different outreach products need to be compiled for different levels of data users; while direct data access is suitable for experts, decision makers need well-condensed policy briefs and journalists often request individual mentoring. By providing these services, the management of the repositories will be professionalized and ready to serve the entire community.

To address these future ambitions, problems of technical equipment and the hiring and long-term retention of qualified personnel must be tackled. Both aspects are required for proper data curation and dissemination of glacier datasets. Hence, in the future GTN-G has to find long-term funding to run all datasets in a mature and sustainable way and serve the community with FAIR and trustworthy glacier data of the best quality.

CONCLUSIONS

Dedicated support and long-term commitment for certified data repositories build the basis for the successful democratization of data. In the field of glacier data, this balancing act has so far been achieved through joint collaboration between data repository institutions, data providers, and data users. From the comparison of seven glacier datasets (**Table 3**) available within the Global Terrestrial Network for Glaciers (GTN-G) we conclude:

- The current state of the GTN-G datasets can largely be explained through historical development, reflecting different needs from stakeholders incl. users.
- The GTN-G datasets have been developed in a research environment, hence long-term data curation and stewardship are absolutely necessary.
- Currently, datasets that are managed based on a mid- to long-term funding (e.g., the FoG dataset) have the highest maturity.
- Urgent action has to be taken to keep the state-of-the-art and individual data curation strategies need to be implemented

and tailored for each operational body, considering the context (e.g., funding situation; project funds vs. long-term funding).

- These strategies need to be evaluated, revised, and adapted on a regular basis, which can be ensured through the GTN-G Advisory Board.
- Data curation requires constant updates of software to meet technical demands of increasing complexity and to provide direct access to most up-to-date information, which in turn needs proper data versioning.
- International standardization such as provided for example by the CoreTrustSeal contributes to a secure setting for the data services.
- Technical equipment, hiring professional staff and long-term retention of qualified personnel is key to offer the different services and to serve the entire community.

Most challenges can only be overcome in a financially safe and secure setting for data services. However, the money spent on the data provider and user side for creating and working with the datasets is several orders of magnitude larger than the funds available for data curation. Considering the importance of glacier data to answer numerous key environmental and societal questions (from water availability to global sea-level rise), this bias needs to be adjusted.

DATA AVAILABILITY STATEMENT

All data are available on the website of the Global Terrestrial Network for Glaciers (GTN-G; www.gtn-g.org) or on the specific websites (see **Table 3**).

AUTHOR CONTRIBUTIONS

IG-R, SN, and MZ conceived the study, assessed the FoG database, assessed the WGI, and assessed the GMC. BR, FP, and MZ assessed the GLIMS database and assessed the RGI. EW, MZ, and IG-R assessed the GlaThiDa. AW, FF, and SN assessed the GPC. IG-R and SN wrote the paper and produced the figures. All authors studied and commented on the selected methodology, reviewed all assessments, and commented on and revised the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This research has been supported by the Federal Office of Meteorology and Climatology MeteoSwiss within the framework of the Global Climate Observing System (GCOS) Switzerland. FP acknowledges additional funding from the ESA project Glaciers_cci (4000127593/19/I-NB). AW and FF acknowledge support from NOAA Cooperative Agreement with CIRES, NA17OAR4320101. BR acknowledges support for Global Land Ice Measurements from Space from National Aeronautics and Space Administration under the National Snow and Ice Data Center and Distributed Active Archive Center.

REFERENCES

- Albani, M., and Maggio, I. (2020). Long time data series and data stewardship reference model, *Big Earth Data* 4, 353–366. doi: 10.1080/20964471.2020.1800893
- Allison, I., Fierz, C., Hock, R., Mackintosh, A., Kaser, G., and Nussbaumer, S. U. (2019). IACS: past, present, and future of the international association of cryospheric sciences. *Hist. Geo Space Sci.* 10, 97–107. doi: 10.5194/hgss-10-97-2019
- Bates, J. J., and Privette, J. L. (2012). A maturity model for assessing the completeness of climate data records, *Eos trans. AGU* 93, 441. doi: 10.1029/2012EO440006
- Bayr, K. J., Hall, D. K., and Kovalick, W. M. (1994). Observations on glaciers in the eastern Austrian Alps using satellite data. *Int. J. Remote Sens.* 15, 1733–1742. doi: 10.1080/01431169408954205
- Bolch, T., Menounos, B., and Wheate, R. (2010). Landsat-based inventory of glaciers in western Canada, 1985–2005. *Remote Sens. Environ.* 114, 127–137. doi: 10.1016/j.rse.2009.08.015
- CEOS. (2020). *WGISS Data Management and Stewardship Maturity Matrix*. Version 1.3, Available online at: https://ceos.org/document_management/Working_Groups/WGISS/Interest_Groups/Data_Stewardship/White_Papers/WGISS%20Data%20Management%20and%20Stewardship%20Maturity%20Matrix.pdf (accessed May 30, 2022).
- Cogley, J. G. (2009). A more complete version of the world glacier inventory. *Ann. Glaciol.* 50, 32–38. doi: 10.3189/172756410790595859
- Farinotti, D., Brinkerhoff, D. J., Clarke, G. K. C., Fürst, J. J., Frey, H., Gantayat, P., et al. (2017). How accurate are estimates of glacier ice thickness? Results from ITMIX, the ice thickness models intercomparison experiment. *The Cryosphere* 11, 949–970. doi: 10.5194/tc-11-949-2017
- Farinotti, D., Huss, M., Fürst, J. J., Landmann, J., Machguth, H., Maussion, F., et al. (2019). A consensus estimate for the ice thickness distribution of all glaciers on Earth. *Nat. Geosci.* 12, 168–173. doi: 10.1038/s41561-019-0300-3
- Field, W. O. (1975). *Mountain Glaciers of the Northern Hemisphere*. CRREL, Hanover, Vol. 1, 698p, Vol. 2, 932p and an Atlas with 49 plates.
- Frey, H., Paul, F., and Strozzi, T. (2012). Compilation of a glacier inventory for the western Himalayas from satellite data: methods, challenges, and results. *Remote Sens. Environ.* 124, 832–843. doi: 10.1016/j.rse.2012.06.020
- Gärtner-Roer, I., Naegeli, K., Huss, M., Knecht, T., Machguth, H., and Zemp, M. (2014). A database of worldwide glacier thickness observations. *Glob. Planet. Change* 122, 330–344. doi: 10.1016/j.gloplacha.2014.09.003
- Gärtner-Roer, I., Nussbaumer, S. U., Hüsler, F., and Zemp, M. (2019). Worldwide assessment of national glacier monitoring and future perspectives. *Mt. Res. Dev.* 39, A1–A11. doi: 10.1659/MRD-JOURNAL-D-19-00021.1
- GlaThiDa Consortium (2020). *Glacier Thickness Database 3.1.0*. Zurich, Switzerland: World Glacier Monitoring Service. doi: 10.5904/wgms-glathida-2020-10
- GLIMS Consortium (2005). *GLIMS Glacier Database, Version 1*. Boulder, CO: NASA National Snow and Ice Data Center Distributed Active Archive Center.
- Guo, W., Liu, S., Xu, J., Wu, L., Shangguan, D., Yao, X., et al. (2015). The second Chinese glacier inventory: data, methods and results. *J. Glaciol.* 61, 357–372. doi: 10.3189/2015JoG14J209
- Hugonnet, R., McNabb, R., Berthier, E., Menounos, B., Nuth, C., Girod, L., et al. (2021). Accelerated global glacier mass loss in the early twenty-first century. *Nature* 59, 726–731. doi: 10.1038/s41586-021-03436-z
- IPCC (2021). “Climate change 2021: the physical science basis,” in *Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, eds V. Masson-Delmotte, P. Zhai, A. Pirani, S. L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M. I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J. B. R. Matthews, T. K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou. Cambridge, UK; New York, NY, USA: Cambridge University Press. doi: 10.1017/9781009157896
- MacGregor, J. A., Studinger, M., Arnold, E., Leuschen, C. J., Rodríguez-Morales, F., and Paden, J. D. (2021). Brief communication: an empirical relation between center frequency and measured thickness for radar sounding of temperate glaciers. *The Cryosphere* 15, 2569–2574. doi: 10.5194/tc-15-2569-2021
- Mercer, J. H. (1967). *Southern Hemisphere Glacier Atlas*. American Geographical Society, US Army Natick Laboratories Technical Report, Natick, 325p.
- Müller, F. (1978). *Instructions for the Compilation and Assemblage of Data for a World Glacier Inventory*. Supplement: Identification/glacier number. Temporary Technical Secretariat for the World Glacier Inventory. Zurich: Swiss Federal Institute of Technology.
- Müller, F., Cafilisch, T., and Müller, G. (1977). *Instructions for Compilation and Assemblage of Data for a World Glacier Inventory*. Temporary Technical Secretariat for the World Glacier Inventory. Zurich: Swiss Federal Institute of Technology.
- National Snow and Ice Data Center (2021). *Glacier Photograph Collection, Version 1*. Boulder, CO: NSIDC: National Snow and Ice Data Center.
- National Snow and Ice Data Center (2021) *Glacier Photograph Collection, Version 1*. Boulder, Colorado USA. NSIDC: National Snow and Ice Data Center. doi: 10.7265/N5/NSIDCGPC-2009-12
- Nussbaumer, S. U., Hoelzle, M., Hüsler, F., Huggel, C., Salzmann, N., and and, M., Zemp (2017). Glacier monitoring and capacity building: important ingredients for sustainable mountain development. *Mt. Res. Dev.* 37, 141–152. doi: 10.1659/MRD-JOURNAL-D-15-00038.1
- Paul, F., Barry, R. G., Cogley, J. G., Frey, H., Haeblerli, W., Ohmura, A., et al. (2009). Recommendations for the compilation of glacier inventory data from digital sources. *Ann. Glaciol.* 50, 119–126. doi: 10.3189/172756410790595778
- Paul, F., and Käab, A. (2005). Perspectives on the production of a glacier inventory from multispectral satellite data in Arctic Canada: Cumberland Peninsula, Baffin Island. *Ann. Glaciol.* 42, 59–66. doi: 10.3189/172756405781813087
- Paul, F., Käab, A., Maisch, M., Kellenberger, T., and Haeblerli, W. (2002). The new remote-sensing-derived Swiss glacier inventory: I. Methods. *Ann. Glaciol.* 34, 355–361. doi: 10.3189/172756402781817941
- Paul, F., Winsvold, S. H., Käab, A., Nagler, T., and Schwaizer, G. (2016). Glacier remote sensing using Sentinel-2. Part II: Mapping glacier extents and surface facies, and comparison to landsat 8. *Remote Sens.* 8, 575, doi: 10.3390/rs8070575
- Peng, G., Privette, J. L., Kearns, E. J., Ritchey, N. A., and Ansari, S. (2015). A unified framework for measuring stewardship practices applied to digital environmental data s. *Data Sci. J.* 13, 231–252. doi: 10.2481/dsj.14-049
- Peng, G., Wright, W., Baddour, O., Lief, C., and the SMMCD Work Group (2019). *The Guidance Booklet on the WMO-Wide Stewardship Maturity Matrix for Climate Data*. Figshare.
- Pfeffer, W. T., Arendt, A. A., Bliss, A., Bolch, T., Cogley, J. G., Gardner, A. S., et al. (2014). The Randolph glacier inventory: a globally complete inventory of glaciers. *J. Glaciol.* 60, 537–552. doi: 10.3189/2014JoG13J176
- Pospiech, M., and Felden, C. (2012). “Big data – A state-of-the-art,” in *AMCIS (Americas Conference on Information Systems) 2012 Proceedings*. Available online at: <https://aisel.aisnet.org/amcis2012/proceedings/DecisionSupport/22>
- Rastner, P., Bolch, T., Mölg, N., Machguth, H., Le Bris, R., and Paul, F. (2012). The first complete inventory of the local glaciers and ice caps on Greenland. *The Cryosphere* 6, 1483–1495. doi: 10.5194/tc-6-1483-2012
- Raup, B., and Khalsa, S. J. S. (2007). *GLIMS Analysis Tutorial. Global Land Ice Measurements from Space (GLIMS)*, 15p. Available online at: https://www.glims.org/MapsAndDocs/assets/GLIMS_Analysis_Tutorial_a4.pdf
- Raup, B., Racoviteanu, A., Khalsa, S. J. S., Helm, C., Armstrong, R., and Arnaud, Y. (2007). The GLIMS geospatial glacier database: a new tool for studying glacier change. *Glob. Planet. Change* 56, 101–110. doi: 10.1016/j.gloplacha.2006.07.018
- Raup, B. H., Kieffer, H. H., Hare, T. M., and Kargel, J. S. (2000). Generation of data acquisition requests for the ASTER satellite instrument for monitoring a globally distributed target: glaciers. *IEEE Trans. Geosci. Remote Sens.* 38, 1105–1112. doi: 10.1109/36.841989
- RGI Consortium (2017). Randolph Glacier Inventory – a data set of global glacier outlines: version 6.0. Technical report. *Global Land Ice Measurements from Space (GLIMS)*, Boulder, CO, Digital Media.
- Scherler, K. E. (1983). *Guidelines for Preliminary Glacier Inventories*. GEMS, UNEP, UNESCO, ICSI, ETH-Z. Temporary Technical Secretariat for the World Glacier Inventory. Zurich: Swiss Federal Institute of Technology.
- UNESCO/IASH. (1970). *Perennial Ice and Snow Masses*. A guide for compilation and assemblage of data for a world inventory. Technical Papers in Hydrology No. 1. Paris: United Nations Educational, Scientific and Cultural Organization.
- Vaughan, D. G., Comiso, J. C., Allison, I., Carrasco, J., Kaser, G., Kwok, R., et al. (2013). *Observations: Cryosphere*. In: *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, eds Stocker, T. F., Qin, D., Plattner,

- G. K., Tignor, M., Allen S. K., Boschung J., Nauels, A., Xia, Y., Bex, V., and Midgley P. M (Cambridge University Press, Cambridge, and New York, NY).
- Welty, E., Zemp, M., Navarro, F., Huss, M., Fürst, J. J., Gärtner-Roer, I., et al. (2020). Worldwide version-controlled database of glacier thickness observations. *Earth Syst. Sci. Data* 12, 3039–3055. doi: 10.5194/essd-12-3039-2020
- WGMS (1989). *World Glacier Inventory – Status 1988*, eds Haeberli, W., Bösch, H., Scherler, K., Østrem, G. and Wallén, C. C. IAHS(ICSU)/UNEP/UNESCO, World Glacier Monitoring Service, Zurich, Switzerland, 458.
- WGMS (1998). Into the second century of worldwide glacier monitoring: prospects and strategies, eds Haeberli, W., Hoelzle, M., and S. Suter. *Studies and Reports in Hydrology*. Paris: UNESCO Publishing, 227.
- WGMS (2018). *Glacier Map Collection (GMC)*. Zurich, Switzerland: World Glacier Monitoring Service. doi: 10.5904/wgms-maps-2018-02
- WGMS (2021a). *Fluctuations of Glaciers Database*. Zurich: World Glacier Monitoring Service.
- WGMS (2021b). Global Glacier Change Bulletin No. 4 (2018-2019). Zemp, M., Nussbaumer, S.U., Gärtner-Roer, I., Bannwart, J., Paul, F., and Hoelzle, M., eds *ISC(WDS)/IUGG(LACS)/UNEP/UNESCO/WMO*. Zurich: World Glacier Monitoring Service, 278.
- WGMS, and National Snow and Ice Data Center (2012). *World Glacier Inventory, Version 1. [Indicate Subset Used]*. Boulder, CO: NSIDC: National Snow and Ice Data Center.
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., et al. (2016). The FAIR guiding principles for scientific data management and stewardship. *Sci. Data* 3, 160018. doi: 10.1038/sdata.2016.18
- Wulder, M. A., Masek, J. G., Cohen, W. B., Loveland, T. R., and Woodcock, C. E. (2012). Opening the archive: how free data has enabled the science and monitoring promise of Landsat. *Remote Sens. Environ.* 122, 2–10. doi: 10.1016/j.rse.2012.01.010
- Zemp, M. (2011). *The Monitoring of Glaciers at Local, Mountain, and Global Scale. Schriftenreihe Physische Geographie; Glaziologie und Geomorphodynamik* 65. University of Zurich, Faculty of Science, 72p.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer LT declared a past collaboration with several of the authors IG-R, SN, FP, and MZ to the handling Editor.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Gärtner-Roer, Nussbaumer, Raup, Paul, Welty, Windnagel, Fetterer and Zemp. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.