Informing Distributed Hydrological Modeling with Worldwide Open Data Services

Strobl, J.* and Nazarkulova, A.
Department of Geoinformatics, University of Salzburg, Austria
E-mail: josef.strobl@plus.ac.at*
**Corresponding Author*

DOI: https://doi.org/10.52939/ijg.v20i12.3763

Abstract

Continuity of water supply is critical for sustaining livelihoods of (Central Asian) mountain communities, which are typically not connected to any regional water utility. Water supply is indispensable for household usage incl sanitation, livestock survival, farming of agricultural crops and small-scale hydropower generation. In mountain environments seasonal as well as multi-annual buffering and storage are limited, surface runoff into streams is the main source of water. The WaterFlow project has been designed to enable academic partners in the Kyrgyz Republic to perform hydrographic analysis of small mountain watersheds to explore current and future potentials for local communities. One key objective is the quantification of the impact of regional climate change scenarios on streamflow as a basis for assessment of future village-level livability and potential mitigation efforts. Hydrologic surface runoff estimation and simulation is a data-intensive process requiring detailed topographic, landcover / landuse and climate change scenario data for distributed modeling. While access to national level or regional data frequently poses a major obstacle to large scale local research. The current and increasing availability of high-resolution global data sets facilitates analyses essentially anywhere on the planet. This paper serves as a case study and process documentation how the 'Living Atlas' as a platform consolidating a rich and continuously updated collection of worldwide openly accessible data supports, inter alia, the foundation data for distributed modeling of surface runoff essentially anywhere on the planet. This research does not aim at achieving results specifically valid for the sample study areas, the latter rather serve as demonstrators for analyses based on cloud-based data services. Steps beyond data preparation, including local observations and integration of climate and land cover change scenarios are briefly mentioned to contextualize the data preparation and pre-processing steps presented below. Altogether, the WaterFlow initiative aims at supporting the development of science-based local capacity for environmental monitoring, within the explicit context of appreciating the potential impacts of likely climate changes on the local water balance.

Keywords: Citizen Science, Geoportals, Hydrological Modelling, Open Data, Watersheds

1. Introduction

Availability of water in several ways is a cross-cutting topic from a Sustainable Development Goals (SDG) perspective and requires support by geospatial data (e.g. topography [1]) and GIScience concepts and methods [2]. Certainly goal #6 "Clean Water and Sanitation" is the SDG most directly addressing access to water [3]. Looking at matters from an agricultural production perspective SDG #2 "Zero Hunger" would have to be mentioned in all dry and arid environments. In mountain regions with a potential for (micro) hydro power contributions to SDG #7 "Affordable and Clean Energy" also is highly relevant. Indirectly SDG #13 "Climate Action" has to be considered from an impact mitigation view. In this particular project with significant elements addressing capacity building, SDG #4 "Quality Education" is supported as well. The wider context of the research documented in this paper is being translated into the more local, small catchment related issues of changes in temporal patterns of streamflow. These impact the sustainability of irrigation agriculture as well as general access to water for all its uses. Assessing current and prospective future change is essential for mitigation measures, and to assess the viability of local municipalities, as most of these mountain communities are dependent on small rivers, which in turn are directly affected by climate change impacts.

ISSN: 1686-6576 (Printed) | ISSN 2673-0014 (Online) |

While the understanding of regional climate scenarios, the dynamics of water-dependent ecosystems and the role of natural risks provide highly important insights, there is a lack of local scale assessment of impacts on livelihoods. Developing an understanding of possible and alternative futures through participation of local actors is considered key to the development of reactive and proactive mitigation steps. The proposed approach is considered a predominantly capacity-building oriented ‘seed initiative’ building the base and creating the potential for a more substantial project aiming at empowerment of local stakeholders through understanding climate-driven environmental dynamics. The development of academic capacities at partner institutions and digitally connected competences of local actors aims at a well-grounded foundation for maintaining livelihoods in a changing environment.

Within the wider context of population dynamics emphasizing urban settings over rural mountain settlements, the viability of the latter not the least depends on reliable and safe access to water. Economic activities based on quality agricultural produce and products as well as tourism offerings can support a local population as long as water supply does not deteriorate to a limiting critical factor. The ‘*Analysis of local changes in water availability for mountain communities’* – ‘WaterFlow’ initiative is pursuing several objectives in different domains, with the overall aim of increasing awareness, capacities and competences towards the management and mitigation of changes in the water cycle in mountain environments.

1. Explore and explain differences in watershed responses to climate scenarios and land cover changes, comparing sample watersheds and identifying critical areas and periods regarding streamflow extremes.
2. Determine the feasibility, data quality and shortcomings of a citizen science / crowdsourcing approach to gauging temporal streamflow patterns.
3. Establish a practice of recording phenological observations as indicators of available moisture.
4. Identify the requirements and establish the feasibility of operating simple atmospheric sensors over annual cycles in online and/or offline mode.
5. Explore the requirements for communicating seasonal observation patterns through a simple digital interface, facilitating individual monitoring and analysis of water balance elements.

2. Sample Study Areas

The watersheds used as sample study areas were jointly identified by project partners. Criteria for selection were (roughly) comparable sizes, accessibility of watershed outlets (pour points) for hydrographic observations, spread across the country and at the same time proximity to partners’ locations, and diversity of elevation ranges, azimuth (exposure), land cover and in particular presence or absence of glaciation. From approx. 25 proposed watersheds the selection was narrowed down to a total of nine, represented in Table 1 and illustrated in Figure 1.

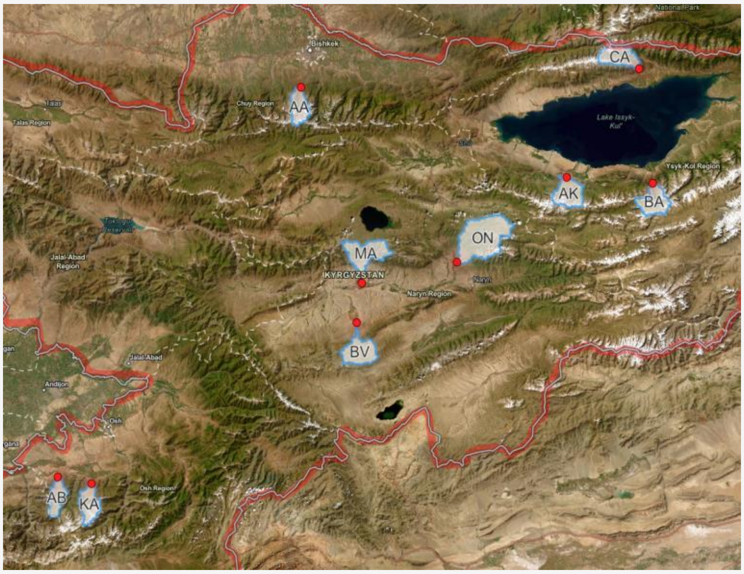


Figure 1: Location and distribution of sample catchments

Table 1: Sample catchments

Code	Catchment name	Area in km²	Pour point locations (E/N, UTM 44N)
AA	Ala-Archa	246.4	-32946 / 4743316
AB	Abshirsay	233.5	-235715 / 4487774
AK	Ak-Say	337.2	155724 / 4669438
BA	Barskoon	319.8	218207 / 4662673
BV	Baetov	378.5	-4189 / 4577811
CA	Chon-Ak-Suu	316.1	211514 / 4741481
KA	Kyrgyz-Ata	302.4	-210838 / 4480608
MA	Moldo-Ashuu	443.1	1547 / 4605148
ON	On-Archa	879.8	72278 / 4615223

Table 2: Living Atlas data services used in WaterFlow

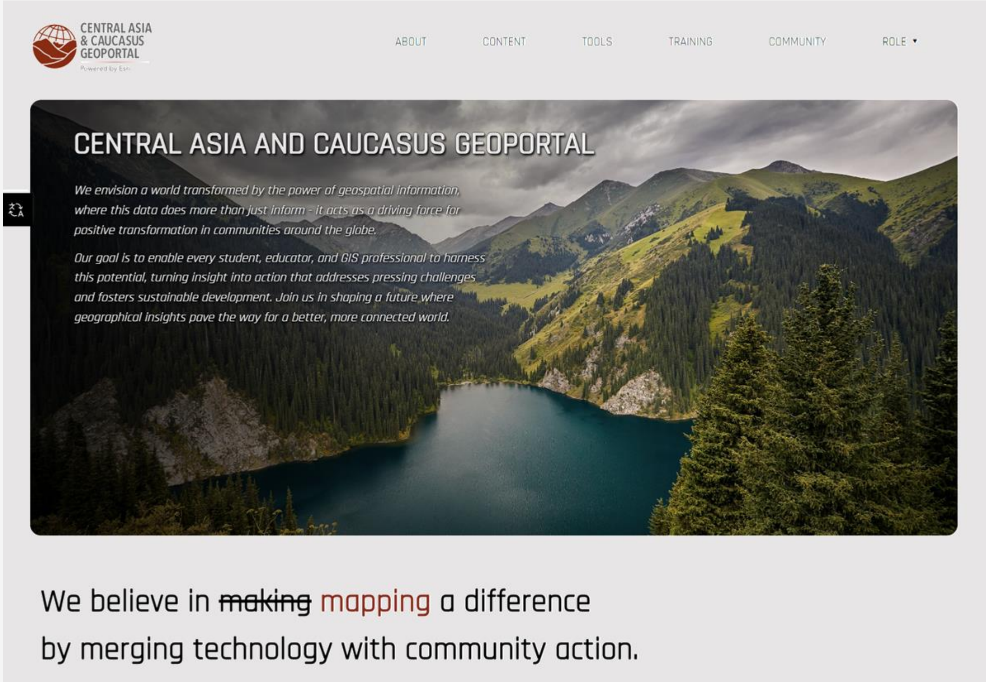
No.	Name	URL	Type
1	Terrain	https://arcg.is/feefK	multi-resolution image service
2	Terrain: Slope in Degrees	https://arcg.is/9fmiv	dynamically calculated image service
3	ESA WorldCover 2021 Land Cover	https://arcg.is/0urHmD	tiled image service
4	Terrain: Hillshade	https://arcg.is/0fXbuf	dynamically calculated image service
5	Terrain 3D	https://arcg.is/1em9Sa	elevation layer (collection)
6	Imagery Hybrid (WGS84)	https://arcg.is/0afbDr	tiled image service
7	World Countries Generalized	https://arcg.is/iGCmW	polygon feature service

All data preparation, calculations and analyses were completed in UTM Zone 44 projection, avoiding distortions and metric inaccuracies resulting from handling in standard Web Mercator. For any resampling steps a NEAREST rule was applied. Delineation of sample watersheds initially was done using the ‘Create Watersheds’ function in ArcGIS Online to create a coarse, generalized outline based on the lower resolution DEM used as a default on this platform. Subsequently this outline was opened in ArcGIS Pro to launch the standard watershed definition workflow [4] for the proper extent. Starting with Fill, followed by Flow Direction (D8) [5], using Flow Accumulation and Snap Pour Point as needed with an adjusted radius to access a local flow accumulation maximum, and finally Watershed-all of these steps based directly on the Terrain service set to UTM Zone 44 and 25m cell size. All 9 resulting watershed rasters then were combined into one by ‘Mosaic to New Raster’ before going through a ‘Raster to Polygon’ conversion. Shared and published as a feature Web Layer to ArcGIS Online, these vectorized watersheds together with their raster counterparts support the following steps for extracting hydrographically relevant layers for each sample catchment.

3. Global Open Data Sources

With the aim of flexibility in the choice of sample study areas and to facilitate worldwide transferability of the chosen approach to distributed hydrologic analyses, the ArcGIS Living Atlas (https://livingatlas.arcgis.com) services are used as the main and nearly exclusive source of data. This includes the resources mentioned above and listed in Table 2. Many of these services are consolidated from different sources, including national mapping agencies, space agencies and other international institutions. Only all their commitments to open data following the FAIR principle facilitates unified accessibility through the Living Atlas portal. The WaterFlow project is basing its educational research case studies on these openly available resources to enable the transferability of its approach to any other study area worldwide.

The selected data sources offer a range of spatial resolutions. Considering typical sizes of catchments (approx. 250 to 800 km²) and maximum resolution of some input data, a common denominator of 25*25m cell size was chosen as a target for resampling and integration of data sets. To provide adequate context for visual interpretation and communication and to avoid cut-out ‘islands’ only showing individual catchments, the Living Atlas also provides seamless basemaps with topographic and imagery styles. These can be selected by viewers and analysts to better understand individual watershed characteristics as well as deciding on next steps in analytical workflows [4] and [6]. Global sources and platforms do not exclude detailed regionalized views, though. As a side note in the context of the work reported here the authors would like to point out access to Living Atlas content plus resources supporting regional and local perspectives in the Central Asia – Caucasus Geoportal (Figure 2).



CENTRAL ASIA AND CAUCASUS GEOPORTAL

We envision a world transformed by the power of geospatial information, where this data does more than just inform - it acts as a driving force for positive transformation in communities around the globe.

Our goal is to enable every student, educator, and GIS professional to harness this potential, turning insight into action that addresses pressing challenges and fosters sustainable development. Join us in shaping a future where geographical insights pave the way for a better, more connected world.

We believe in ~~making~~ mapping a difference
by merging technology with community action.

Figure 2: Central Asia - Caucasus Geoportal at https://www.cacgeoportal.com

This initiative serves as a community platform for geospatial mapping, analysis and communication across all scientific and societal domains, welcoming everyone to join and benefit from this well organized and rich collection of data.

4. Data Preparation for Sample Catchments

For distributed runoff modeling (and to a much lesser degree for lumped area parameter models) the key factors for infiltration and evapotranspiration, runoff pathways, flow velocity, retention etc. need to be derived from digital elevation models, soil, and land cover data and of course an understanding of current surface conditions. For catchment-scale analyses the spatial distribution of these factors needs to be known, below key workflow steps are outlined:

Elevation values were extracted from the multiresolution Terrain service using the ArcGIS RECLASSIFY function with 100m intervals, masked by watershed and resampled with a NEAREST rule to 25 by 25m resolution. Class frequencies are coded with the lower range threshold, e.g., 2700 for the 2700m to <2800m class and extracted from the layer attribute tables.

Slope values were extracted from the Terrain: Slope in Degrees service dynamically calculated from Terrain elevation values resampled with a NEAREST rule at 25 by 25m resolution. The RECLASSIFY function with 5° intervals and masking by watershed creates the desired results. Class frequencies again are coded with the lower range threshold, e.g., 35 for the 35° to <40° class. It must be noted that slope values are systematically underestimated compared to a (potentially) higher resolution DEM.

Similarly, land cover [7] as a key element in retention of precipitation, evapotranspiration and runoff modification needs to be considered. In this case study, the European Space Agency WorldCover 2021 Land Cover [8] data are used for characterization of catchments. Watershed masks are applied with 10m resampling (corresponding to the source resolution) through the ‘Extract by Mask’ function. All the above-mentioned factors (elevation, slope, land cover) are statistically summarized by watershed and fed into dashboard visualization as table feature services. To support visualization at larger scales, a streams layer has been generated with the ‘Derive Stream As Line’ function from the hydro-conditioned DEM with an upstream threshold value of 1 km². These flow networks are clipped to exact catchment boundaries and visualized according to the Strahler stream order [9] and [10]. Further inputs for runoff modeling parametrization can be generated with the same approach, either extracting data like other land cover or simulated land cover change by watershed extent from Living Atlas services, or using hosted algorithms to generate e.g., landform classes from geomorphons [11] by on-demand server-side processing (Figure 3).

5. Exploration Dashboard

For initial exploration by researchers and to support generating hypotheses, basic catchment characteristics are presented in an interactive dashboard (Figure 4) allowing inspection of morphological as well as land cover metrics through a combination of map centric as well as aggregate statistics views. This approach has the advantage of enabling researchers to explore multiple parameters quickly and effectively and to compare hydrographically relevant watershed characteristics. The main orientation dashboard view allows the selection of one of the nine sample watersheds which is then highlighted on the map. At the same a hypsometric curve, slope function and a pie chart with land cover class proportions is displayed. For each diagram, the underlying data series can be inspected by activating the 'Table' tab.

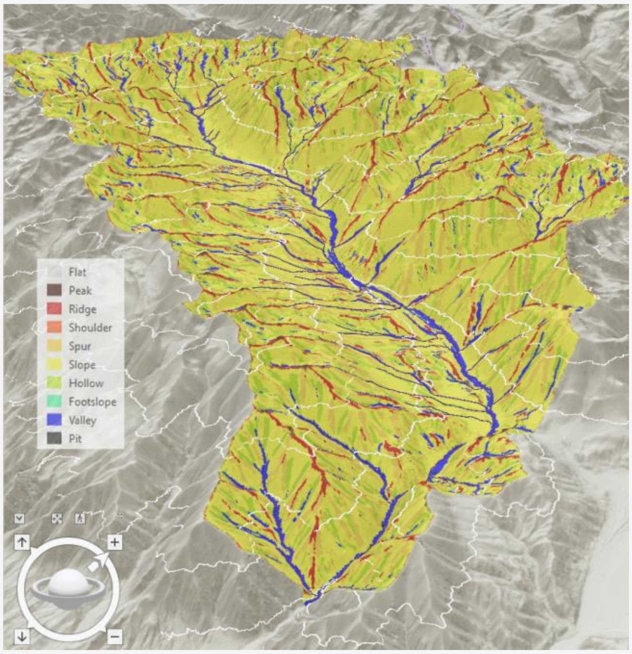


Figure 3: Landform classes derived from DEM-based geomorphons

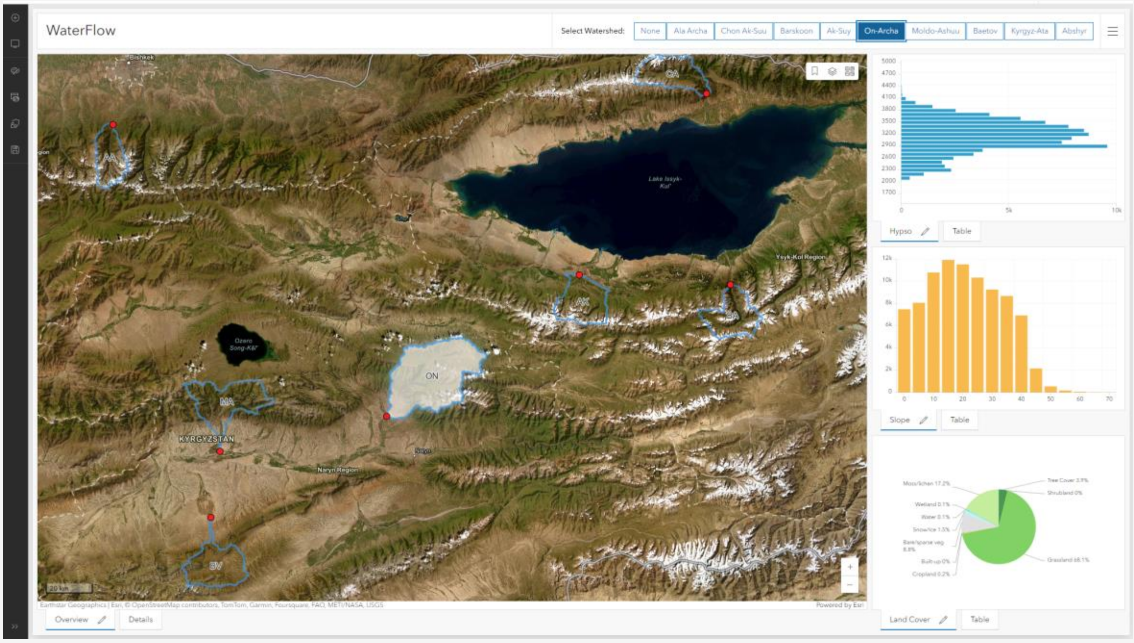


Figure 4: Dashboard overview

Grassland

This class includes any geographic area dominated by natural herbaceous plants (Plants without persistent stem or shoots above ground and lacking definite firm structure): (grasslands, prairies, steppes, savannahs, pastures) with a cover of 10% or more, irrespective of different human and/or animal activities, such as: grazing, selective fire management etc. Woody plants (trees and/or shrubs) can be present assuming their cover is less than 10%. It may also contain uncultivated cropland areas (without harvest/ bare soil period) in the reference year

Figure 5: Dashboard detail view, focus on one sample watershed

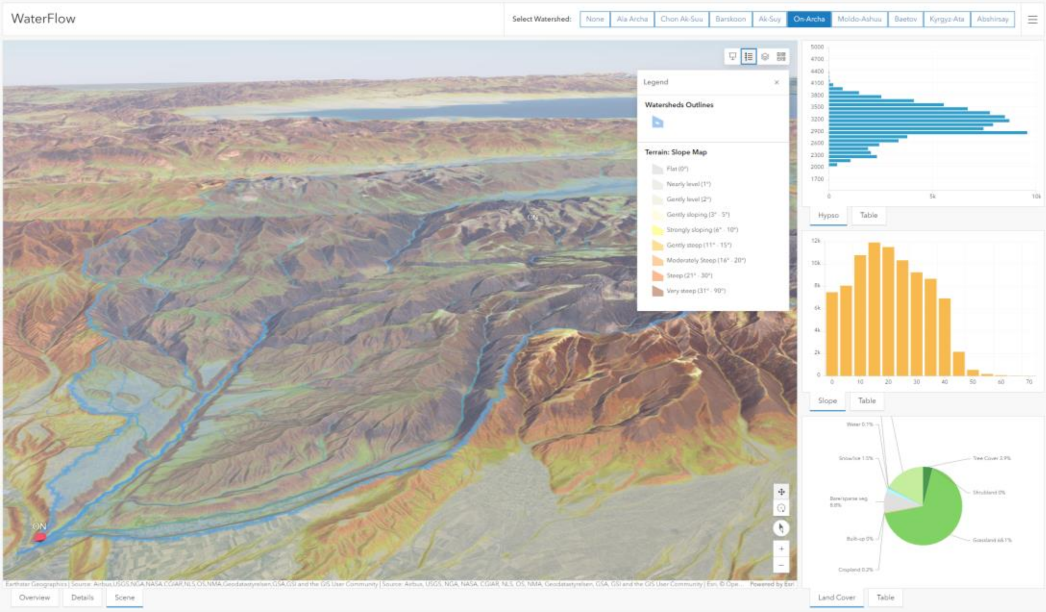


Figure 6: Freely navigable perspective view with choice of slope classes

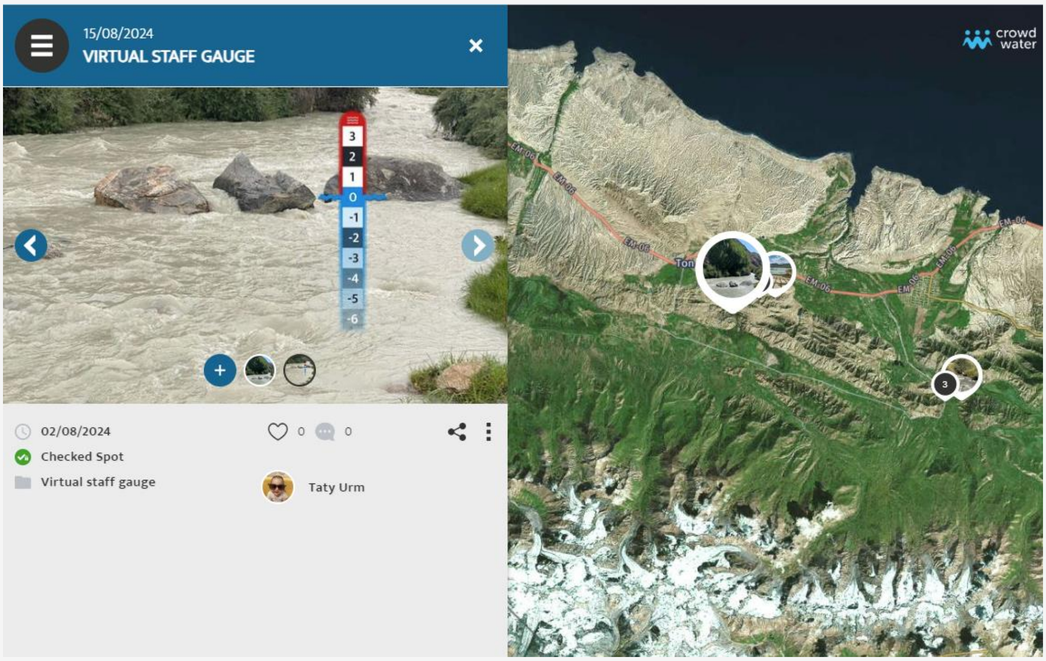
The map area in the dashboard is fully navigable and allows activation of a variety of map layers, of course including elevation, slope, and land cover as well as additional supporting elements. This and other dashboards are openly accessible, experts therefore can share the URL with other stakeholders. While the user experience design is not oriented towards interaction with the general public, researchers can introduce actors in specific regions to the functionality (and conceptual background) and allow 'anyone' to use this platform for individual exploration. Whenever 'diving deeper' into one of the sample watersheds is required, this is managed by switching with the Details tab to a localized view and pan/zooming to the target area (Figure 5). This will support more in-depth exploration of data sets by selecting layers, exploring these through popups while again having access to statistical distributions through charts. In addition, flow lines are displayed automatically when exceeding a set scale factor for better topographic orientation. Perspective views enable a better understanding of local topographies and their impact on morphographic factors as well as the distribution of landscape elements within a watershed (Figure 6).

By activating the Scene tab and choosing a watershed with an initial default view angle through a bookmark slide, analysts are enabled to freely navigate perspective and scale and are given the option to select from a range of layers for display. Based on all these approaches to building familiarity with the ‘lay of the land’ and the distribution of relevant runoff impact factors analysts are given the opportunity to take more informed approaches to runoff modeling. While the latter is outside the scope of this paper, the authors believe in the power of exploratory uses of geovisualization to enhance the setup of advanced modeling and simulation (not only) in hydrology.

6. Dynamic Online Services

While the above demonstrated data services are fundamental for any kind of distributed hydrological analysis, modelling and simulation it is essential to understand that surface runoff processes cannot be represented by deterministic approaches, and that the temporal dimension is equally important as the spatial domain. Local observations of runoff patterns therefore are indispensable for calibrating and for temporally enabling the analysis of processes. The identified sample watersheds thus are intended to be monitored for runoff patterns. Since standard hydrographic stations including runoff gauges are expensive to install and maintain and consequently concentrated along a few major streams (or in preparation of hydroengineering projects), a more modest approach is chosen for the WaterFlow initiative. Project partners currently are using a well-researched geospatial Citizen Science [12] and [13] approach (evolving into participatory science) aimed at developing time series of runoff to better understand individual watershed dynamics. The CrowdWater service platform [14] and [15] facilitates ongoing observations at watershed pour points, and as web service is ‘live linked’ to web maps and results are accessible through live web maps or dashboards (see Figure 7). Another example for real-time data streams is the ECMWF GEOGLOWS [16] medium term forecasting service, linked into the above presented interfaces. Display and query can be optionally activated by users, again demonstrating the full integration capabilities of map services with temporally oriented series of observations and dynamic forecast data streams.

These elements are only briefly mentioned here to indicate that the value and importance of working with live online services in the context of a topical spatial analysis domain is not restricted to accessing worldwide map and feature services, but also enables the integration with dynamic streams of observations. This of course also would include not only the output (outflow) dimension of watersheds and runoff models, but also the input dimension of meteorological point observations as well as remotely sensed precipitation, atmospheric and surface condition data – all these elements are not further explored at this point.



15/08/2024
VIRTUAL STAFF GAUGE

02/08/2024
Checked Spot
Virtual staff gauge
Taty Urm

Figure 7: CrowdWater virtual gauge at pour point of Ak-Say watershed

7. Summary and Perspectives

The example workflows and exploratory interfaces presented in this paper demonstrate the capabilities of current global geospatial portals for analytical applications anywhere, worldwide. With distributed hydrographic modeling of small area watersheds as the target domain, all required base data for subsequent runoff modeling are prepared by exclusively using services available on the Living Atlas platform. While spatial analytical applications in earlier years more often than not were constrained by limited availability of data, Geoinformatics methods have moved from a scarcity of data to ubiquitous and pervasive supplies. Access to digital representations of real-world features and phenomena is not a limiting factor anymore, analysts therefore now can focus on the extraction of information from data and feeding these into models to answer important 'what-if' questions across domains.

Having access to global datasets with worldwide coverage at high spatial and temporal resolution satisfying the needs of most applications also enables the transferability of concepts, analytical workflows, models, and insights towards other regions on the planet. Results are not only achievable for specific study areas where extensive databases have been laboriously compiled but can be obtained for any other region of interest simply by shifting the spatial extent. Certainly not all limitations have been removed yet: spatial resolutions better than the 10-30m cell range are not available with worldwide coverage, major short-term (e.g., land cover) changes in less-than-one-year timeframes might not be represented and not all ontological contexts are fully provided for. Nonetheless, the limiting factor today rather is the human capacity to ask the right questions and the competence to apply algorithms and methods towards intended insights [17].

Of course, open access to global geospatial data also serves as the key foundation to move forward with AI approaches, in many cases already shortcutting the human-centric workflow from data to interpretation and insights. While these developments obviously are at the center of attention today, the current rate of progress only has become possible due to the move from data paucity to a deluge of geospatial data. Whether this rapid technical rate of advancements keeps up with the rate of change in our environments and societies and with the evolution of competences and 'brainware' [18] is an entirely different question.

Acknowledgements

The WaterFlow project is supported through the Austrian 'Kooperation Entwicklungsforschung' (KoEF) program with funding provided by the Federal Ministry for Education, Science and Research (BMBWF) and managed by the Austrian Agency for Education and Internationalization (OeAD). The authors express their gratitude to partners in the WaterFlow initiative actively contributing to fieldwork, specifically Ermek Baibagishov, Akylbek Chymyrov, Gulshaan Ergeshova, Nursultan Ismailov and Tatygul Urmambetova.

References

[1] Xiong L., Li, S., Tang, G. and Strobl, G., (2022). Geomorphometry and Terrain Analysis: Data, Methods, Platforms and Applications. *Earth-Science Reviews*, Vol. 233, https://doi.org/10.1016/j.earscirev.2022.104191.

[2] Scott, G. and Rajabifard, A., (2017). Sustainable Development and Geospatial Information: A Strategic Framework for Integrating a Global Policy Agenda into National Geospatial Capabilities. *Geo-spatial Information Science*, Vol. 20(2), 59-76. https://doi.org/10.1080/10095020.2017.1325594.

[3] Baskaran, V. and Velkennedy, R., (2022). A Systematic Review on the Role of Geographical Information Systems in Monitoring and Achieving Sustainable Development Goal 6: Clean Water and Sanitation. *Sustainable Development*, Vol. 30(5), 1417-1425. https://doi.org/10.1002/sd.2302.

[4] Maidment, D., (1993). *Handbook of Hydrology*. McGraw Hill.

[5] Tarboton, D. G., (1997). A New Method for the Determination of Flow Directions and Upslope Areas in Grid Digital Elevation Models. *Water Resources Research*, Vol. 33(2), 309-319. https://doi.org/10.1029/96WR03137.

[6] Maidment, D. and Morehouse, S., (2002), *ArcHydro: GIS for Water Resources*. Esri Press.

[7] Strobl, J. and Nazarkulova, A., (2022). Land Cover Cloud Analytics: from Global Services to Regional Insights. *International Journal of Geoinformatics*, Vol. 18(6), 1–9. https://doi.org/10.52939/ijg.v18i6.2451.

[8] Zanaga, D., Van De Kerchove, R., Daems, D., De Keersmaecker, W., Brockmann, C., Kirches, G., Wevers, J., Cartus, O., Santoro, M., Fritz, S., Lesiv, M., Herold, M., Tsendbazar, N. E., Xu, P., Ramoino, F. and Arino, O., (2022). *ESA WorldCover* 10 m 2021 v200 [Data set]. *Zenodo*, https://doi.org/10.5281/zenodo.7254221.

[9] Tarboton, D. G., Bras, R. L. and Rodriguez-Iturbe, I., (1991). On the Extraction of Channel Networks from Digital Elevation Data. *Hydrological Processes*, Vol. 5, 81–100. https://doi.org/10.1002/hyp.3360050107.

[10] Metz, M., Mitasova, H. and Harmon, R. S., (2011). Efficient Extraction of Drainage Networks from Massive, Radar-Based Elevation Models with Least Cost Path Search. *Hydrology and Earth System Sciences*, Vol. 15(2), 667-678. https://doi.org/10.5194/hess-15-667-2011.

[11] Jasiewicz, J. and Stepinski, T. J., (2013). Geomorphons-A Pattern Recognition Approach to Classification and Mapping of Landforms. *Geomorphology*, Vol. 182, 147-56. https://doi.org/10.1016/j.geomorph.2012.11.005.

[12] Goodchild, M. F., (2007). Citizens as Sensors: The World of Volunteered Geography. *GeoJournal*, Vol. 69(4), 211–221. https://doi.org/10.1007/s10708-007-9111-y.

[13] Rickles, P., Haklay, M., Ellul, C. and Skarlatidou, A., (2017). Citizen Science with GIS&T. *The Geographic Information Science & Technology Body of Knowledge* (3rd Quarter 2017 Edition), John P. Wilson (ed.). https://doi.org/10.22224/gistbok/2017.3.5.

[14] Etter, S., Strobl, B., van Meerveld, I. and Seibert, J., (2020). Quality and Timing of Crowd-Based Water Level Class Observations. *Hydrological Processes*, Vol. 34(2), 4365-4378. https://doi.org/10.1002/hyp.13864.

[15] Wang, Z., Seibert, J., van Meerveld, I., Lyu, H. and Zhang, C., (2023). Automatic Water-Level Class Estimation from Repeated Crowd-Based Photos of Streams. *Hydrological Sciences Journal*, Vol. 68(13), 1826-1840. https://doi.org/10.1080/02626667.2023.2240312.

[16] Hales, R., Nelson, J., Souffront, M., Guttierez, A., Prudhomme, C., Kopp, S., Ames, D., Williams, G. and Jones, N., (2022) Advancing Global Hydrologic Modeling with the GEOGloWS ECMWF Streamflow Service. *Journal of Flood Risk Management*, https://doi.org/10.1111/jfr3.12859.

[17] Nazarkulova, A. and Strobl, J., (2023). Digital Earth Competences Across Disciplines. *International Journal of Geoinformatics*, Vol. 19(11), 20–25. https://doi.org/10.52939/ijg.v19i11.2917.

[18] Strobl, J., (2019). The Geospatial Capacity Building Ecosystem - Developing the Brainware for SDI. *In book: Sustainable Development Goals Connectivity Dilemma. Chapter: 13*. CRC Press - Taylor & Francis. https://doi.org/10.1201/9780429290626.