





RESEARCH ARTICLE

Building a unified sustainable development goal database: Why does sustainable development goal data selection matter?

Anne Warchold¹  | Prajal Pradhan¹  | Pratibha Thapa^{1,2} |
 Muhammad Panji Islam Fajar Putra^{1,3}  | Jürgen P. Kropp^{1,4} 

¹Potsdam Institute for Climate Impact Research (PIK), Member of Leibniz Association, Potsdam, Germany

²Technische Universität Bergakademie Freiberg, Faculty of Business Administration, Freiberg, Germany

³Carbon Disclosure Project (CDP) Europe, Berlin, Germany

⁴Institute for Environmental Science and Geography, University of Potsdam, Potsdam, Germany

Correspondence

Prajal Pradhan, Potsdam Institute for Climate Impact Research (PIK), Member of Leibniz Association, Potsdam, Germany.
 Email: pradhan@pik-potsdam.de

Funding information

German Federal Ministry of Education and Research for the BIOCLIMAPATHS project (grant agreement no. 01LS1906A) under the Axis-ERANET call.

Abstract

The 2020s are an essential decade for achieving the 2030 Agenda and its Sustainable Development Goals (SDGs). For this, SDG research needs to provide evidence that can be translated into concrete actions. However, studies use different SDG data, resulting in incomparable findings. Researchers primarily use SDG databases provided by the United Nations (UN), the World Bank Group (WBG), and the Bertelsmann Stiftung & Sustainable Development Solutions Network (BE-SDSN). We compile these databases into one unified SDG database and examine the effects of the data selection on our understanding of SDG interactions. Among the databases, we observed more different than similar SDG interactions. Differences in synergies and trade-offs mainly occur for SDGs that are environmentally oriented. Due to the increased data availability, the unified SDG database offers a more nuanced and reliable view of SDG interactions. Thus, the SDG data selection may lead to diverse findings, fostering actions that might neglect or exacerbate trade-offs.

KEYWORDS

inequalities, SDG indicator databases, SDG interactions, SDG networks, SDGs, synergies and trade-offs

1 | INTRODUCTION

The 2020s are a critical decade for achieving the 2030 Agenda and thus for transforming our world towards a more sustainable one. To measure the agenda's progress, the United Nations (UN) adopted 17 Sustainable Development Goals (SDGs) and 169 targets in 2015. Five years into its implementation, however, no country is on track in meeting the SDGs until 2030 (Sachs et al., 2020). Additionally, the COVID-19 pandemic has also negatively impacted most SDGs (Pradhan et al., 2021). The global SDG research community continues to address this challenge by finding evidence-based guidance for policymakers and stakeholders to develop and implement the strategies needed to achieve the SDGs. To foster exchanges between

science, policy, and society, SDG research needs to provide sound evidence to ensure that scientific outcomes are translated into concrete and practical actions.

Currently, SDG-related publications have reached a staggering 4.1 million articles, presenting both opportunities and challenges for global SDG research (Elsevier, 2020). Science-based approaches in SDG research range from qualitative, static-, to dynamic-quantitative methods applied at different scopes, ranging from sector- or goal-specific to integrated ones and from local to global scales (Allen et al., 2021). Since SDGs are a system of interacting components rather than just a collection of goals, targets, and indicators (Pradhan, 2019), SDG research, however, needs to go beyond one approach and scale. Within the SDG system, synergies and trade-offs

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *Sustainable Development* published by ERP Environment and John Wiley & Sons Ltd.

need to be disentangled to support decision-making and prioritizing actions. Proving this support requires, for example, the identification of positive or negative multiplication effects within the system (Pham-Truffert et al., 2020), of nonlinear SDG interactions that can trigger rapid progress with minimal investments (Warchold et al., 2021), or of entry points to support the SDG system, such as climate initiatives (Coenen et al., 2021). Consequently, SDG decision-making should not just be normative but also evidence-based to ensure efficient resource utilization. In this sense, data holds the potential to support and inform actions toward realizing the SDGs, but also great potential to marginalize or misinform about SDGs' progress.

Quantitative SDG research primarily uses databases provided by the United Nations, the World Bank Group (WBG), and the Bertelsmann Stiftung & Sustainable Development Solutions Network (BE-SDSN) at a global scale. For example, Pradhan et al. (2017), Warchold et al. (2021), and Anderson et al. (2021) used the UN data for a holistic quantification of SDG interactions. Kroll et al. (2021) and Asadikia et al. (2021) statistically investigated SDG interactions using data from the BE-SDSN. Lusseau and Mancini (2019) and Laumann et al. (2020) estimated the system of SDG interactions using the WBG data. Further, some SDG research does not apply any of those databases but instead simplifies the complexity of the SDGs by using at least one indicator per SDGs based on different sources or own model results (Jägermeyr et al., 2017; Mainali et al., 2018; Obersteiner et al., 2016; Zhang et al., 2022). Comparing the findings of those studies show similar but also different results. For example, Pradhan et al. (2017) detect SDG 8 and 12 being the goals most commonly associated with trade-offs. Contrarily, the results by Laumann et al. (2020) evince that economic growth, and by Asadikia et al. (2021) that responsible consumption, do not play such a central role for sustainable development compared to other SDGs. Miola and Schiltz (2019) compared existing SDGs performance tools using the same set of indicators for EU member states. They detected differences in the tools, where the selection of indicators and methods applied led to substantially different relative evaluations of the SDGs. We go beyond evaluating single SDG performances and apply the same statistical method to measure SDG interactions holistically using different databases. The question, therefore, arises as to whether the data selection is the cause of the differences in results, which would make it difficult to compare SDG-related articles and their policy implications. This aspect already emphasizes the importance of a unified SDG framework and database to obtain comparable results (Miola & Schiltz, 2019; Warchold et al., 2021).

To assess how the selection of SDG data matters, we apply a temporal correlation analysis to investigate synergies and trade-offs within and across SDGs at the global, income, and regional scales, and evaluate the corresponding results. Further, this study provides the first framework to unify SDG databases. Based on these analyses, we showed the strengths and limitations of each of the databases and similarities and differences among them. Additionally, we investigated variations among the databases by developing simple SDG networks to identify the most connected goals and targets based on the results from the correlation analyses. Both approaches enable the

identification of SDG goals and targets that are synergistic or impeding in the achievement of the 2030 Agenda.

2 | DATA AND METHODOLOGY

2.1 | SDG monitoring frameworks and databases

Indicators monitor progress, develop implementation strategies, and manage resource allocations for achieving the 2030 Agenda. The UN, the WBG, and the BE-SDSN currently provide global SDG frameworks and corresponding databases widely used by scientists, policymakers, and practitioners. Those SDG frameworks try to cover the multidimensional aspects of sustainable development and are translated into SDG indicators. Here, we provided a brief description of these databases, which are elaborated in Text S1.

Together with national statistical offices, the Inter-agency and Expert Group on SDG Indicators (IAEG-SDGs) develops methodologies and compiles data for SDG indicators, which they submit to the UN Statistics to generate the *Global SDG Indicators database* (United Nations Statistics Division, 2020). We refer to this database as the UN database. Currently, the UN database provides data on 192 SDG indicators for a total of 258 countries and areas between 1967 and 2020 (download August 20, 2020) (United Nations Statistics Division, 2020). For almost every indicator, the UN database provides sub-indicators, which further are disaggregated. This data disaggregation considers demographic factors (e.g., gender, age group, or rural-urban) but also nondemographic factors (e.g., cities, sectors, or products) (Data S1). We refer to those disaggregated UN SDG sub-indicators as UN indicator data.

The WBG extracted relevant indicators from their premier data compilation, the *World Development Indicators* (WDI), and reorganized them according to the goals and targets of the 2030 Agenda. We will refer to it as the WBG database. Several of the WBG indicators are also disaggregated by demographic and nondemographic factors. To get a comparable foundation and terminology, we will refer to the WBG disaggregated indicators as WBG indicator data. Currently, the WBG database offers 371 indicator data for a total of 215 countries and areas between 1990 and 2019 (download August 6, 2020) (World Bank Group, 2020a).

The BE-SDSN also outlined how a comprehensive SDG indicator framework and associated monitoring systems might be established to support the goals and targets of the 2030 Agenda (Sustainable Development Solutions Network, 2015). The BE-SDSN provide open access SDG data, which we refer to as the BE-SDSN database. These indicators, however, are only mapped to the 2030 Agenda at the goal level. The BE-SDSN framework consists of 114 indicators with 85 global indicators and 29 indicators added specifically for OECD countries. Despite none of those 114 indicators being disaggregated, we use the same terminology as for the UN and WBG and refer to them as BE-SDSN indicator data. The published BE-SDSN database from 2020 offered data for 85 indicator data for a total of 193 countries between the years 2000 and 2020 (download August 3, 2020) (Sachs et al., 2020).

2.2 | Establishment of unified SDG database

We conducted three steps to merge the UN, WBG, and BE-SDSN databases into one unified SDG database (abbreviated unified database). First, we consider the lowest common denominator of years—2000 to 2019—as a comparison period. The Millennium Development Goals (MDGs), which served as the data foundation for several SDGs, were adopted in 2000. Therefore, we choose 2000 as the first year for the database.

In a second step, we assigned the UN, WBG, and BE-SDSN indicator data to the officially adopted global SDG indicator framework, that is, the 17 SDGs and 169 targets. Since the UN database is fully built around that framework, no adjustments were needed. The WBG database consists of indicator data, which are already assigned to the target level. We took over those target assignments and allocated WBG indicator data to either similar UN indicator data or added them as additional indicators to cover the target. The allocation of WBG to UN indicator data occurs if both indicator data had the same or showed close resemblances in the description. If the description was similar, but the unit differs, we assigned both indicator data to cover the target. We particularly focused on disaggregated data while merging the UN and WBG frameworks. Despite the UN disaggregating data more frequently, disaggregated WBG data still supplement the unified SDG framework. To merge the BE-SDSN indicator data to the unified SDG framework, we first individually assigned all 85 indicators to the SDG target level. Subsequently, we decided per target whether the BE-SDSN indicator data is similar or resembles already allocated UN or WBG indicator data. If all three showed similar descriptions with the same meaning and the same unit, we allocated them as one indicator data to cover the target within the unified SDG framework. However, most BE-SDSN indicators do not overlap with the indicators of the other databases but present additional measurements to cover SDG targets and supplement the unified SDG framework.

In the third step, we decided on a unique list of indicator data for the unified database (Data S1). We reviewed each of the 169 targets and the assigned indicator data from the UN, WBG, and BE-SDSN to maximize the number of usable indicators. If indicator data were overlapping (Table S1), meaning UN, WBG, or BE-SDSN provide data for the indicator, we choose the database with the highest data availability over time and space. Our criteria for the data availability is the number of data points—the amount of available values for the respective indicator for all countries over the years (Data S1 lists the exact amount of data points per indicator data). If the amount of data points were the same, we preferred the UN data.

2.3 | Statistical analysis of SDG interactions at different scales

For the comparison of the SDG interactions, our methodological approach is twofold. First, we use the statistical method introduced by Pradhan et al. (2017) to explore synergies and trade-offs among SDGs based on (anti)correlations. Accordingly, we apply a temporal analysis, measuring correlations between a pair of indicator data for

each country. We measure the Spearman's ranked correlation coefficient (ρ) between at least eight paired observations (representing indicator data values for one country for at least 8 years) (Interstate Technology & Regulatory Council, 2013). These indicator data pairs can belong to the same SDG or two distinct SDGs. Some indicator data are measured by so-called dummy variables (also known as the Boolean indicator or binary variable). Those indicators consist of numerical values either being 0 or 1 to indicate the absence or presence of categorical effects. We exclude these indicator data from our analysis because of the bias for the statistical analysis.

The coefficient value ρ is multiplied by the relation's direction: a positive sign refers to indicators that are desirable to increase and a negative sign to those indicators that need a decline for meeting the 2030 Agenda. Based on the resulting coefficient's sign and value, we define synergies and trade-offs. A plus sign indicates a positive relation (synergy), and a minus sign indicates a negative one (trade-off). To avoid over-interpretation of correlation (Hauke & Kossowski, 2011), we implement thresholds while defining synergies and trade-offs. We define SDG interactions with $\rho \in (0.5, 1]$ as “synergies,” with $\rho \in [-0.5, 0.5]$ as “not-classified,” and with $\rho \in [-1, -0.5]$ as “trade-offs” (Smarandache, 2016; Warchold et al., 2021).

For comparing the UN, WBG, BE-SDSN, and unified databases, we analyze SDG interactions at the global, income, and regional scales. The World Bank Atlas categorizes countries based on their income into low-income countries (LIC), lower-middle-income countries (LMIC), upper-middle-income countries (UMIC) and high-income countries (HIC) (Figure S1) (World Bank Group, 2020b). We use this grouping to investigate how a country's macroeconomic context influence SDG interactions based on different databases. Since SDG interactions could also vary due to different country's social and environmental factors, we investigate SDG interactions based on world regions. We summarize countries into four world regions based on the United Nations Regional Groups (United Nations, 2020): Western World, Latin America, Asia-Pacific, and Africa (Figure S2).

By capturing SDG synergies and trade-offs at global, income, and regional scales, we further distinguish similarities and differences among the SDG databases. At the global scale, we distinguish the top 10 SDG pairs with similarities and differences in SDG interactions among SDG databases. Considering income groups and regions, the four databases do not provide consistent data for the 153 SDG pairs (binomial coefficient of 17 SDGs, Table S2). For a reasonable comparison of similarities and differences across the income-based and regional SDG interactions, we only choose SDG pairs having indicator data across all four databases. The amount of available SDG pairs for the income and regional groups varies significantly across the databases. Due to the reduced data availability and, therefore, comparable SDG pairs at income and regional scales, we only distinguish the top five SDG pairs with similarities and differences in SDG interactions among the SDG databases. In terms of similarities, we define an SDG pair as a top pair if it exhibits the highest shares of synergies or trade-offs for all four SDG databases. For differences, we consider an SDG pair as a top pair when it has the highest range of synergies (R_S) or trade-offs (R_T) across the databases, that is, $R_S > 50\%$ or $R_T > 30\%$.

Availability/databases ^a	UN	WBG	BE-SDSN	Unified
Time span (least common denominator)	2000–2019			
Countries (and areas)	248	215	193	255
Goals	17	17	17	17
Targets	135	75	49 ^b	143
Indicators	181	–	–	–
Disaggregation level	29	9	–	23
Total indicator data ^c	2317	366	85	2584

Note: The 17 goals, 169 targets, and 243 indicators are according to the Interagency and Expert Group on SDG indicators (IAEG-SDGs) official global SDG indicator framework.

^aThe SDG data for the UN was downloaded on August 20, 2020, for the WBG on August 6, 2020, and for the BE-SDSN on August 8, 2020.

^bWe assigned the BE-SDSN indicators to related SDG targets.

^cWe referred to disaggregated SDG indicators as indicator data.

TABLE 1 Sustainable Development Goals (SDGs) data availability provided by the United Nations (UN) (United Nations Statistics Division, 2020), the World Bank Group (WBG) (World Bank Group, 2020a), the Bertelsmann Stiftung & Sustainable Development Solutions Network (BE-SDSN) (Sachs et al., 2020) in August 2020 in comparison to a unified database

Second, we apply network analysis based on the correlation analysis results, namely the share of synergistic and impeding SDG interactions instead of the absolute values. The method adopted in this paper is an extension of the approach proposed by Putra et al. (2020) and Lusseau and Mancini (2019). We generate the network of interactions at two scales: among the 169 SDG targets and the 17 SDGs. By creating networks, we identify the most positively and negatively connected goals and targets in the SDG system for each of the four databases. Based on these networks, we can visualize the interactions and assess the role of components within the system. In general, a network structure consists of a set of nodes, either goals or targets and a set of edges, which are the connection between the nodes based on the correlation.

To simplify the network structure and focus more on the significance of nodes in the networks, we assigned three conditions for defining goal and target networks. First, we excluded target connections with a low amount of indicator data pairs (<20) to cover more connected targets in the networks and to represent global phenomena. Since the goal network aggregates the target network, we considered all goal connections. Second, we apply thresholds for the edges representing synergistic or impeding interactions between the nodes. We did not include edges with shares (s) in synergies (s_s) or trade-offs (s_t) with $s \leq 30\%$. This enables us to distinguish structurally connected nodes within the goal and target networks. Third, we calculate the eigenvector centrality of nodes to assess their significance in the network based on the transitive influence of nodes. A high eigenvector value indicates that a node is connected to many nodes, which themselves have high scores.

3 | RESULTS

3.1 | Comparison of SDG databases

The three SDG databases have their strengths and limitation in covering the 2030 Agenda (Table 1). Although there are similarities between the databases, the degree of differences outweighs them,

and thus they complement each other. However, complete indicator data time-series are not available for all time steps and countries. The UN database provides data for the highest number of countries and areas over time. However, the data availability varies significantly across regions and time. The same holds for the WBG database. To minimize biases from missing data, the BE-SDSN only includes countries that have data for at least 80% of the indicators across time (Sachs et al., 2020). For that reason, the BE-SDSN database provides the lowest number of countries but offers the most consistent time-series.

The amount of indicator data per target and goal varies significantly across the three databases (Table S3, Data S1). BE-SDSN use a mix of official and nonofficial data sources, including model-based estimates, to fill data gaps and reduce time lags in official statistics. Therefore, all BE-SDSN indicator data have an established methodology and good data coverage over time and space. The data coverage tends to be better for socioeconomic goals such as SDG 2, 3, 4, and 9, whereas data availability to monitor SDG 10 and 13 is relatively low. Despite defining a few SDG 12 indicators, BE-SDSN provides no downloadable data for this goal. However, BE-SDSN contributes some unique indicator data on environmental aspects (SDG 13–15). For example, the BE-SDSN database measures energy-related CO₂ emissions and those embodied in imports (Target 13.2). The WBG database has similar characteristics. Here, SDGs related to socio-economic aspects have the largest number of indicators, like SDG 4. SDG 13 and 14 have the smallest number of indicator data. Even so, the WBG database provides a unique environmental focus on, for example, overfishing (Target 14.4) and biodiversity (Target 15.5).

Compared to the WBG and BE-SDSN databases, the UN database's amount of indicator data is many times larger, covering two to three times as many targets. Although SDG 17 has the highest number of targets, the amount of indicator data is largest for SDG 4. One reason for the high number of UN indicator data is the high level of disaggregation within the SDGs. By contrast, SDG 7 and 13 have the lowest number of targets and the smallest number of UN indicator data. Still, the UN database considers 29 different levels of

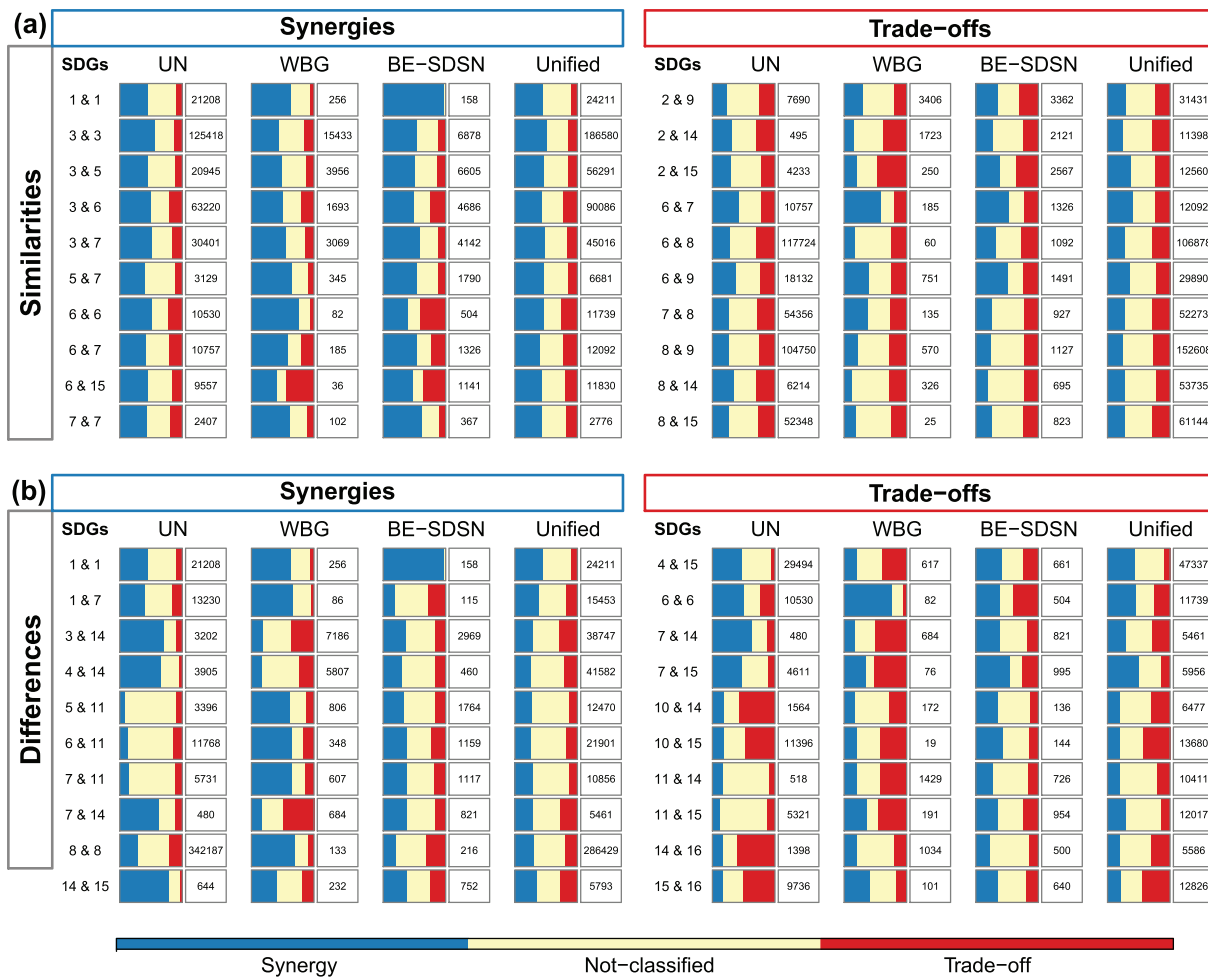


FIGURE 1 Top ten Sustainable Development Goal (SDG) pairs with similarities (a) and differences (b) in interactions based on the four databases—the United Nations (UN), World Bank Group (WBG), Bertelsmann Stiftung & Sustainable Development Solutions Network (BE-SDSN), and unified—globally. The color bar represents the synergies (blue), not-classifieds (yellow) and trade-offs (red). The numbers in the boxes represents the number of data pairs used for the analysis. The numbers in left hand side of the figures represents the SDG pairs [Colour figure can be viewed at wileyonlinelibrary.com]

disaggregation and provides the most distinct view on demographic and nondemographic factors. The WBG database offers nine diverse levels of disaggregation with some overlaps with the UN database but also consists of unique disaggregation into specific types of species (Target 15.5), industries (Target 6.4, 8.2), and target values (Target 11.6).

Our unified database offers more indicator data, covering more targets across more countries than the other three SDG databases (Tables 1, S2–S4, and Data S1). The unified database consists of a unique list of 2584 indicator data for 255 countries and areas between 2000 and 2019. Even if the indicator data increase does not seem much, the amount of available data pairs did increase drastically, especially for income and regional groups (Table S1). Our database provides a comparable number of indicator data per target with a more consistent data disaggregation. Still, this database provides no data for 26 targets, mainly associated with SDG 11 to 17 (Data S1).ⁱ Both, the unified SDG framework (Data S1) and the unified SDG database (Data S2), are provided in the Supporting Information.

3.2 | Global SDG interactions

At the global scale, we observe similarities and differences in SDG interactions while considering the four databases (Figures S3–S6, Data S3 and S4). On average, synergistic interactions always outweigh impeding ones regardless of the data selection (Table S5). Instead of explaining details on interactions for each SDG pair, we highlight the top 10 SDG pairs with similarities and differences in SDG synergies and trade-offs among the four databases (Figure 1). Mainly, we compare the interactions based on the UN, WBG, and BE-SDSN databases, as they show more extreme variations.

The top 10 SDG pairs with similarities in synergies and trade-offs have some resemblances. These pairs have similar shares of synergies or trade-offs among all databases, that is, shares of synergies greater than 40% or trade-offs greater than 20%. We observe similarities, particularly for those SDGs that are social- and economically-oriented, with a higher number of indicator data, and that have, therefore, more consistent data over time and space. Within these top 10 pairs with

similarities, SDG 3, 6, and 7 appear four times each (Figure 1A, left) due to similar indicator data for these goals across all four SDG databases (Data S1). For example, all databases consist of indicators on mortality rates and diseases (Target 3.1, 3.2, 3.3, 3.4), safely managed drinking water and sanitation and freshwater withdrawal (Target 6.1, 6.2, 6.4), and access to clean energy, renewable energy shares, and energy intensity (Target 7.1, 7.2, 7.3). Interactions within SDG 3, 6, and 7 exhibit broad compatibility of indicators, where one indicator's progress is associated with the fulfillment of another one in the same goal. We also observe the inter-goal synergistic interactions of these SDGs reflected by them recurring within the top 10 SDG pairs. Energy access (Target 7.1) powers medical facilities (Target 3.1–3.4) and strengthens water treatments and distribution systems (Target 6.1, 6.2) (UN-Water, 2016). This improved access to safe water and hygiene services (Target 6.1, 6.2) is beneficial for homes, healthcare facilities, and schools, which directly support a number of targets on nutrition and health (Target 2.2, 3.1–3.3, 3.9), education (Target 4.1–4.3), and gender equality (Target 4.5, 5.2, 5.5). Reducing inequality in energy- and water-related infrastructures (SDG 6, 7) and access to healthcare (SDG 3) leads to synergies with other goals, regardless of data selection.

Contrarily, SDG 8 appears in five of the top 10 SDG pairs with similarities in trade-offs (Figure 1A, right). All four databases provide similar indicator data on (un-)employment rates (Target 8.2, 8.3, 8.5, 8.6) and on bank account ownership (Target 8.10). The top impeding SDG interactions among all four SDG databases is between SDG 6 and 8, with shares between 20% and 30%. For example, the sustainment of per capita economic growth (Target 8.2) impedes the water-use efficiency and freshwater withdrawal (Target 6.4) through excessive use of resources (UN-Water, 2016) and increases adverse environmental impacts (SDG 14 and 15). However, an adequate and reliable supply of water (Target 6.1, 6.4) is essential for many economic activities (SDG 8). Further, targets related to increasing economic or agricultural productivity, industrialization, expansion of infrastructure, and urbanization (SDG 2, 7–9), if not sustainably implemented, negatively impact the use of natural resources and therewith primarily results in trade-offs with environmental-related targets (SDG 6, 14, and 15). These recurring trade-offs among SDG 2, 6, 7, 8, 9, 14, and 15 in all four databases reflect a chain of impeding interactions independent of the data selection. These interactions emphasize that achieving sustainable management and efficient use of natural resources requires transformation on how energy and manufacturing systems use natural resources to minimize their adverse impacts (Fuso Nerini et al., 2018).

We further depict the top 10 SDG pairs with differences in synergies and trade-offs across all four databases globally (Figure 1B). The share of synergies ranges from 43% to 53% among the databases in the top 10 SDG pairs with differences in synergies. We observe the highest difference in synergistic shares between the UN and the WBG database. Results obtained from the WBG database show high synergistic shares for SDG pairs 5 and 11; 6 and 11; and 7 and 11, whereas for the UN database, those are not-classified. Ensuring access to safe and affordable housing (Target 11.1) can increase security and safety, especially for women (Target 5.1–5.5), improve access

to adequate sanitation and clean drinking water (Target 6.1, 6.2), as well as access to electricity and clean fuels and technologies for cooking (Target 7.1). The WBG database not only consists of indicators on the population living in slums (like the UN) but also on the urban population. We observe high shares of synergies for these indicators with the above-mentioned targets. Additionally, the WBG, in comparison to the UN, provides detailed data on the proportion of population exposed to air pollution (Target 11.6). A primary risk factor for deaths is the use of solid fuels for cooking (Target 7.1), which causes indoor air pollution (Target 11.6). To avoid this form of pollution, households should switch from traditional cooking and heating methods towards more modern and cleaner practices (Ritchie & Roser, 2019).

Contrarily, results obtained from the UN database show high synergistic shares for SDG pairs 3 and 14; 4 and 14; 7 and 14; and 14 and 15, whereas for the WBG database, those are rather not-classified or exhibit higher shares in trade-offs. The WBG database only consists of indicators for Target 14.4 and 14.5, while the UN database offers data for Target 14.3, 14.5, 14.6, 14.a, and 14.b. For example, we detect high synergistic interactions between an increasing allocation of renewable resources (Target 7.2) and reduced marine acidity level (Target 14.3) based on UN data. Ocean acidity has increased about 25% from preindustrial times to the early 21st century, a pace faster than any known in Earth's geologic past (U.S. Environmental Protection Agency, 2019). The global adoption of renewable energies to power human activity reduces emissions of carbon dioxide and other acid-forming compounds and, therefore, contributes to acidification (U.S. Environmental Protection Agency, 2019).

The interaction between SDG 1 and 1 is in the top ten pairs for similarities (approx. $s_S > 45\%$ across all databases) and differences in synergies. This is because the range in synergies (R_S) simultaneously exceeds due to the high shares of synergies based on the BE-SDSN database ($s_S > 97\%$) our threshold of detecting differences in synergies ($R_S > 50\%$). The same logic applies for SDG pair 6 and 6 with similarities in synergies but also differences in trade-offs. Among the top 10 SDG pairs with differences in impeding interactions, the shares of trade-offs range between 32% and 50%. We observe these large differences for goals related to lives below water (SDG 14) and on land (SDG 15) (Figure 1B, right). One reason behind this observation is the diverse indicator data for SDG 14 and 15, which also holds for SDG 10 and 11 (Data S1). SDG pairs 7 and 14, 7 and 15, 11 and 14, and 11 and 15 indicate impeding interactions based on the WBG database, which are rather not-classified for other databases. Energy systems can impact marine and terrestrial ecosystems. For example, unsustainably managed energy services can drive local pollution, ocean acidification, ecosystem loss and degradation associated with the use of fuelwood, or competition for space with energy infrastructure (Fuso Nerini et al., 2018). The SDG pairs 10 and 14, 10 and 15, 14 and 16, and 15 and 16 exhibit trade-offs with shares greater than 50% based on the UN data but significantly lower shares based on the other databases. Despite biodiversity on land and oceans bears a tremendous potential to mitigate climate change and benefits our society and economy (Cardinale et al., 2012), we detect many trade-offs in regards to SDG 14 and 15. This finding

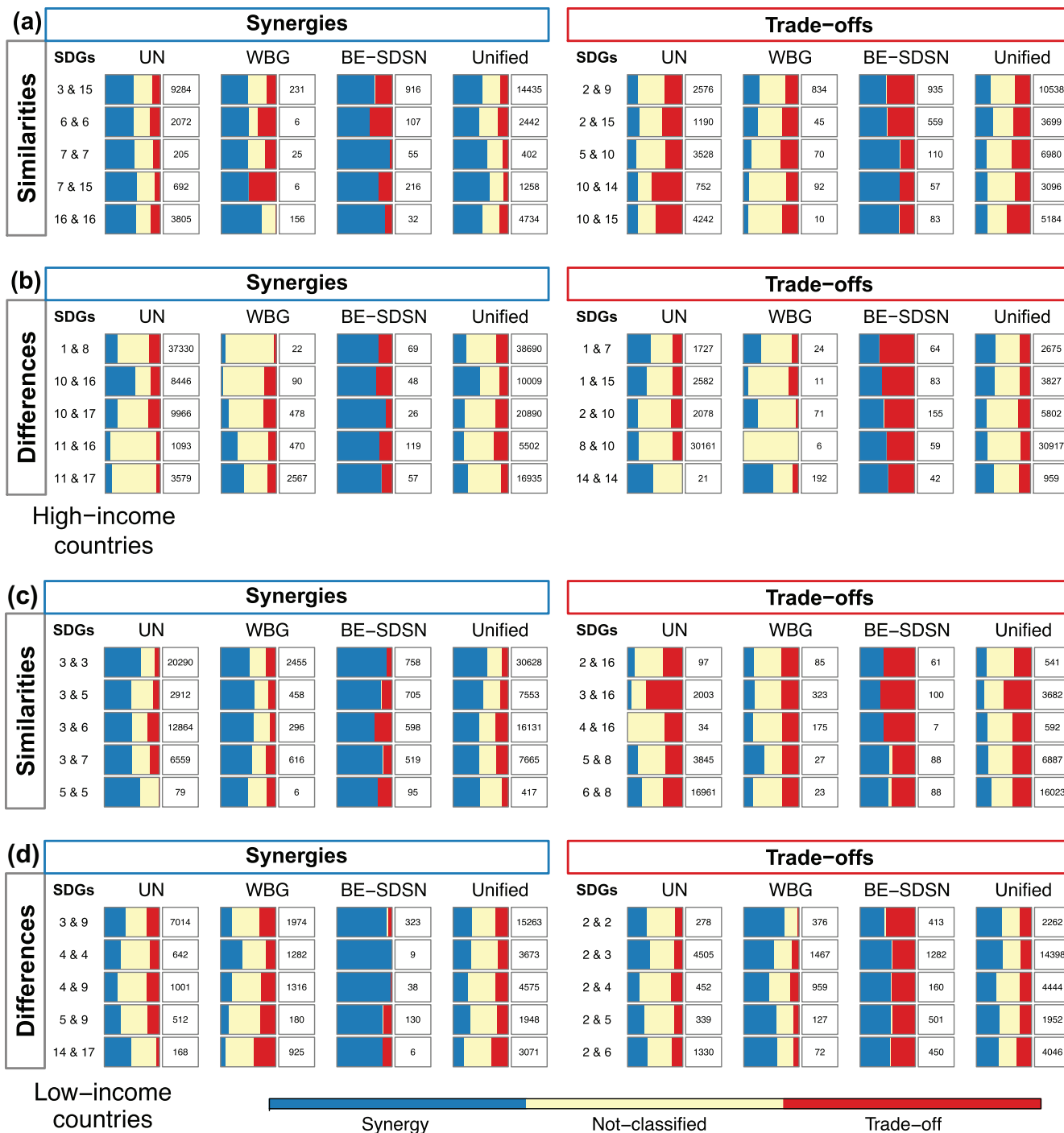


FIGURE 2 Top five Sustainable Development Goal (SDG) pairs with similarities (a,c) and differences (b,d) in interactions based on the four databases—the United Nations (UN), World Bank Group (WBG), Bertelsmann Stiftung & Sustainable Development Solutions Network (BE-SDSN), and unified—for high-income (a,b) and low-income (c,d) countries. The color bar represents the synergies (blue), not-classifieds (yellow) and trade-offs (red). The numbers in the boxes represents the number of data pairs used for the analysis. The numbers in left-hand side of the figures represents the SDG pairs [Colour figure can be viewed at wileyonlinelibrary.com]

highlights that the current development path negatively impacts biodiversity in the name of societal well-beings and economic growth. Therefore, biodiversity-centered solutions should be increasingly integrated into global sustainable development strategies to exploit underlying synergies with other goals to transform the current development path.

3.3 | SDG interactions across income groups

Within the four income groups, SDG interactions vary significantly, comparing the four SDG databases (Figure 2). Independent of the data selection, synergies on average outweigh trade-offs in each income

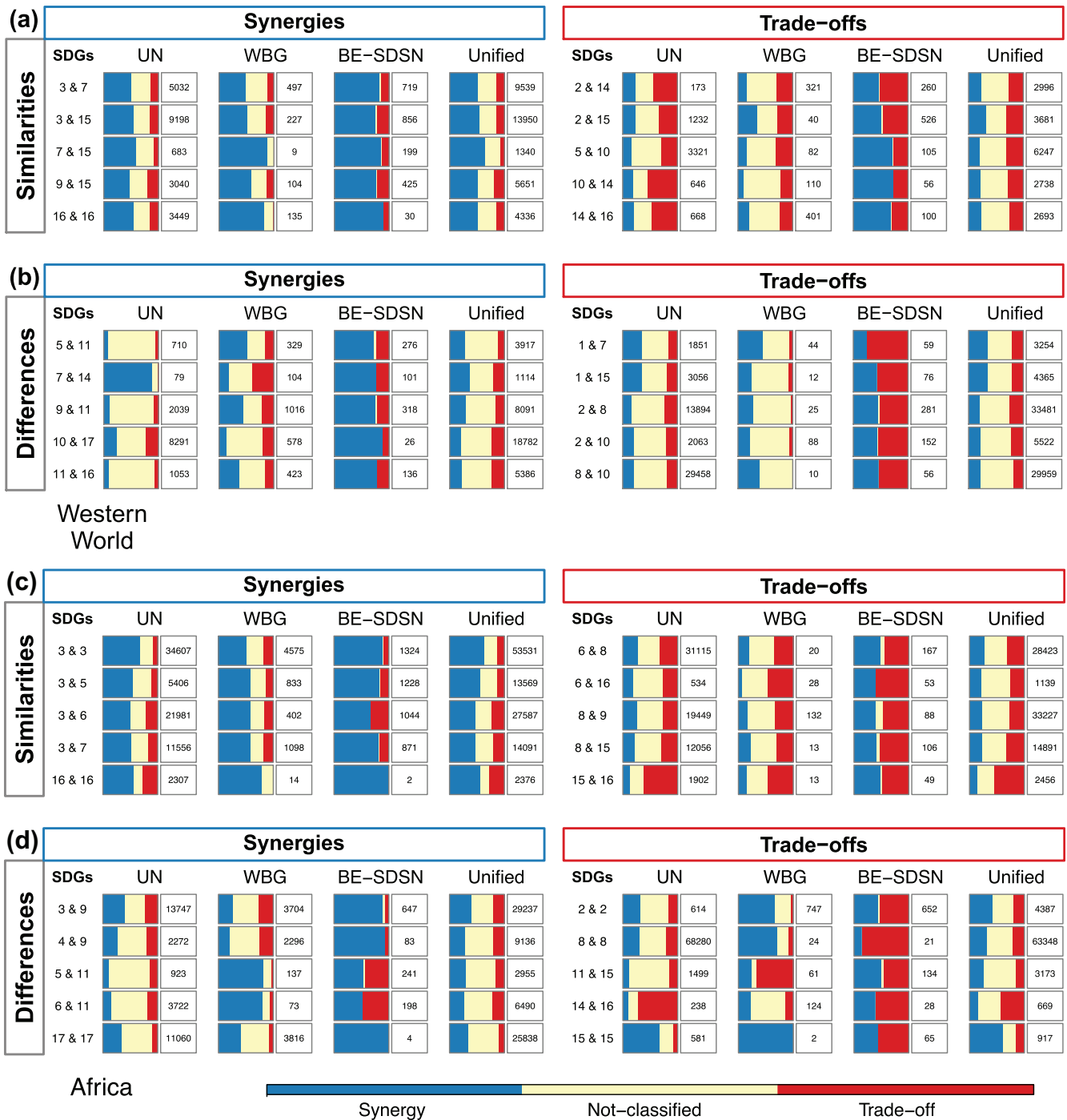


FIGURE 3 Top five Sustainable Development Goal (SDG) pairs with similarities (a,c) and differences (b,d) in interactions based on the four databases—the United Nations (UN), World Bank Group (WBG), Bertelsmann Stiftung & Sustainable Development Solutions Network (BE-SDSN), and unified—for the Western world (a,b) and Asia-Pacific (c,d). The color bar represents the synergies (blue), not-classifieds (yellow) and trade-offs (red). The numbers in the boxes represents the number of data pairs used for the analysis. The numbers in left-hand side of the figures represents the SDG pairs [Colour figure can be viewed at wileyonlinelibrary.com]

group within and across SDGs (Table S5). Likewise, some SDG interactions show a similar share in synergies or trade-offs within each income group for all databases (Figure 2A, C). For example, SDG 6, 7, and 16 have synergistic intra-goal interactions in HICs and UMICs, and SDG 8, 10, and 16 impeding inter-goal interactions. Within LICs

and LMICs, SDG 3 reoccurs in the top five SDG pairs with synergies. The similar share of indicator data for SDG 3, 6, and 7 across all four SDG databases is one reason behind this observation. Independent on the data selection, SDG 8 has negatively connections to other SDGs in most income groups (except for HICs). These trade-offs are related to

the tensions within the income groups to rapidly expand access to essential basic services, the need for efficient energy systems, and to ensure economic growth, causing impeding interaction with other SDGs.

Despite those similarities, income groups also have differences in shares of synergistic and impeding interactions among the four SDG databases (Figure 2B,D for LICs and HICs, Figure S7B,D for LMICs and UMICs). For example, LICs show the lowest share of synergies based on the UN database (average sum of 30.4%), followed by the unified and WBG database (average sum of 30.5% and 38.1%, respectively), and the highest shares based on the BE-SDSN database (average sum of 59.4%) (Table S3). This order of databases does not necessarily apply to the other income groups. For LICs, SDG 9 reoccurs in three of the top five SDG pairs with differences in synergies. According to our target assignment of indicator data, the BE-SDSN database only provides data for targets 9.5 and 9.c (Data S1). Enhancing scientific research (Target 9.5), particularly in developing countries, is founded on supporting quality in early childhood education (SDG 4). Significantly increasing access to information communications technology (ICT) (Target 9.c) addresses key health system challenges in LMICs and LICs. ICT-enabled health programs, for instance, extend geographic access to health care, improve data management, therewith, improve diagnosis and treatment, and mitigate fraud and abuse (SDG 3 and 5) (Lewis et al., 2012). Since the UN, WBG, and unified database offer additional SDG 9 indicator data, their interactions are also not-classified or impeding.

The database selected for analysis SDG interactions related to hunger, nutrition, and food security (SDG 2) in LICs can likewise lead to contrary findings. Where inter-goal interactions of SDG 2 based on the BE-SDSN database show high shares of synergies and trade-offs, the WBG data-based results exhibit mainly synergies and the UN data-based results not-classified SDG 2 interactions. Selecting the unified database leads to an almost even distribution of shares, despite synergies still outweighing trade-offs. These differences relate to the use of different indicator data, the varying amount of indicator data pairs (Figure 2, numbers in the boxes) and SDG pairs (Table S1), consequently the smaller number of indicator data representing an SDG pair (especially on the part of the WBG and BE-SDSN), and the lower data availability, missing data or data inconsistency over time or space. The unified database shows fewer extreme deviations in shares of synergies or trade-offs due to the higher data volume, creating a more nuanced and robust picture of SDG interactions within income groups.

If the level of income is a descriptor of SDG interactions, the share of synergies and trade-offs should either be more significant within income groups than globally or show trends with increasing income levels. Results based on the BE-SDSN database exhibit higher shares of synergies and trade-offs within the income groups than globally. Comparing the other three databases to the global level, however, the results are diverse. These varying shares of synergies and trade-offs not just occur due to income inequalities but also because of data selection. However, especially the substantial inequalities between income groups need to be tackled to achieve the 2030 Agenda.

3.4 | SDG interactions across regions

Assessing SDG interactions for regions in relation to the data selection reveals some resemblances to the income groups' results. Each region has on average more synergistic than impeding SDG interactions regardless of the data selection (Table S5). However, the extent of shares differs for each region depending on the data selection. For example, Asia-Pacific shows the lowest share of synergies based on the WBG database (average sum of 25.6%), followed by the unified and UN database (average sum of 30.6% and 32.5%, respectively), and the highest shares based on the BE-SDSN database (average sum of 53.9%).

To highlight how SDG interactions vary, we focus on the top five SDG pairs with similarities and differences in synergies and trade-offs among the four SDG databases (Figure 3—the Western World and Asia-Pacific, and Figure S8—Africa and Latin-America). For all four regional groups, SDG pair 16 and 16 has high shares of synergies among all SDG databases. Those intra-goal synergies imply a robust result independent of the data selection. However, the SDG 16 indicators cover a wide range of issues including homicide, violence, prisoners, corruption and bribery indicators. Those indicators across different issues are, therefore, in themselves strong correlates.

For the Western World, the top five SDG pairs in terms of synergies consist of SDG 15 and trade-offs of SDG 14, independent of the data selection. In case of trade-offs with SDG 2, a biodiversity loss (SDG 14 and 15) could be related to the growth of intensive crop cultivation and industrial livestock production (SDG 2) (Ladha et al., 2020; Tsiafouli et al., 2015). Although the databases do not use identical indicators for SDG 2, 14, and 15, they still refer to the same problem that needs to be addressed for sustainable transformation. For Africa, the top five SDG pairs with synergies consist of SDG 3, 5, 6, and 7 independent of the data selection. SDG 3 is prominent for synergies among all databases. SDG 8 occurs in three of the top five pairs with similarities in trade-offs. These results resemble those in LICs, wherefore, the same explanation for the manifestation of similarities applies. Those resemblances in results between the regions and income groups arise since the Western World mainly consists of HICs and Africa of LICs (Figures S1 and S2). Similar analogies emerge when comparing other regional with income groups' results (e.g., UMICs with Latin America or Asia-Pacific).

We further detect differences in SDG interactions based on the data selection. Our results for Africa present another analogy compared to results of LICs, namely differences in synergies for SDG pairs 3 and 9 and 4 and 9. Additionally, SDG 11 reoccurs in the top five SDG pairs with differences in synergies for the Western World and Africa. In the case of Africa, these differences arise to the extent that the SDG 11 interactions based on the UN database are not-classified, based on the WBG database are synergistic, based on the BE-SDSN database show a 50-50 distribution of synergies and trade-offs, and based on the unified database a 25-50-25 distribution of shares. SDG 11 has different indicators across the SDG databases (Data S1), with a large variation in the number of provided data (Table S3). The

UN database offers 264 indicator data for Target 11.1, 11.5, 11.6, and 11.b. The WBG database provides nine indicator data (Target 11.1, 11.6) and the BE-SDSN database, four indicator data (Target 11.1, 11.2, 11.6). This finding highlights that the data selection significantly defines city-related SDG interactions.

The top five SDG pairs of Africa, Latin America, and Asia-Pacific with differences in trade-offs repeatedly contain SDG 14 and 15. Selecting the UN, WBG, or BE-SDSN database to assess how human activities affect biodiversity results in large shares of trade-offs for different SDG interactions within all three regional groups. Only the unified database provides a distribution of shares that is scaled between the extremes based on the other three databases. Similar to the given explanation at the global scale, the SDG 14 and 15 indicator data is quite diverse across the databases. Additionally, the data availability and consistency throughout the biodiversity-related SDG pairs vary within the regions depending on the data selection.

3.5 | SDG networks

At goal and target levels, the structure of the SDG networks changes with the data selection (Figures S9–S12 for goals, Figures 4, 5, S13, and S14 for targets). These networks enable identifying the most positively and negatively connected SDG goals and targets in all databases. In general, SDG networks at the target level scale up to the goal level because we observe more positive than negative connections within the networks based on all databases. Nevertheless, these networks also represent several instances of negative connections at the goal and target levels.

In the synergistic goal networks, the most connected SDG varies according to the selected database. We observe the highest eigenvector centrality for SDG 14 in the UN-based, SDG 7 in the WBG-based, and SDG 3, 9, and 15 in the BE-SDSN-based goal network (Figures S9–S11A, Table S6). For the unified database, SDG 3 and 15 have the largest eigenvector centrality (Figure S12A, Table S6). Although the synergistic networks are more complex, the trade-offs are not negligible as they could negatively impact the achievement of the 2030 Agenda. Especially, the fulfillment of the environmentally related goals (SDG 12–15) could be hindered by progress in other SDGs (Figures S9B–S12B), as those have large eigenvector centralities, especially in the UN, WBG and unified-based goal networks (Table S6). Although the most negatively connected goals depend on the data selection, the networks display that achieving environmental-related SDGs will be challenging if current trends continue.

Since the goal networks aggregate the results of the target networks, both show similar but also different results. However, the most positively or negatively connected target does not necessarily belong to the most positively or negatively connected goal. Most SDG goal interactions exhibit synergies or trade-offs with shares of $30% < s < 45%$, whereas at the target level, we observe more strong connections ($30% < s \leq 100%$) (Data S3 and S4). The number of connections per node of goal networks does not necessarily reflect the one in the

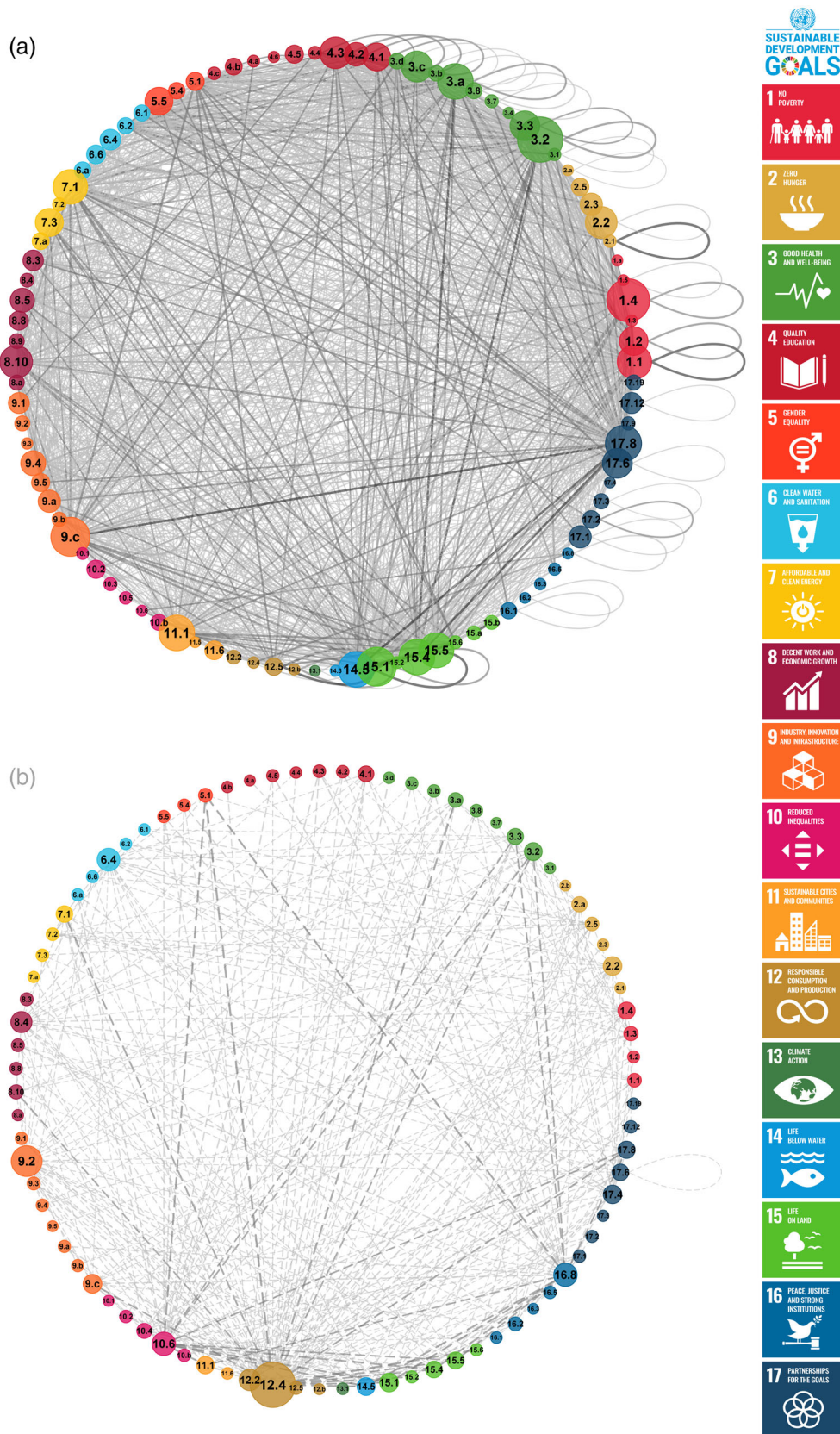
target networks. This aspect especially applies to the impeding networks, where, for example, the goal network based on the unified database shows the least connections per node. In contrast, the target network based on the unified database has the most connections per node. For these reasons, it is essential to gain more nuanced insights into interactions at the target level, as the share of synergies and trade-offs can more prominently be distinguished.

At the target level, the networks are more complex to the extent that they display more nodes, more positive than negative connections, and more diverse shares of synergies and trade-offs. The target networks based on the unified database (Figure 4) display more nodes with connections showing synergies (97 targets) and trade-offs (87 targets). In comparison, the WBG has approximately 50% fewer nodes and connections (49 targets with synergies and 43 targets with trade-offs, Figure 5).

Independent of the data selection, the target networks reveals that reducing child mortality (Target 3.2) is a clear structural component in the SDG system (Figures 4, 5, S13, and S14). Target 3.2 has the largest eigenvector centrality in all networks (Table S7), reflecting that the interactions with other targets are multipliers of the networks and will positively affect many other target interactions. Further, Target 11.1 (ensure access for all to adequate, safe and affordable housing and basic services and upgrade slums) exert much positive connection to other targets in the network based on the WBG, but also to some extent unified and BE-SDSN database. However, despite Target 3.2 and 11.1 having the most positive interactions with other targets, the connections have synergistic shares of $30% < s \leq 60%$. Our findings show that the biodiversity-related targets, i.e., Target 14.5, 15.1, 15.4, and 15.6, are not only strongly connected to the achievement of other targets but also strongly synergize with each other ($s > 90%$). These results can be found within the unified, UN, and BE-SDSN data-based target networks. Conversely, Target 14.4 and 15.1 were found to have the most negative connections to targets of other SDGs in the network based on the WBG database (Figure 5B). Despite the positive intra-goal target connections, our results suggest that achieving biodiversity-related goals may also face obstacles from progress in other targets. Nonetheless, these conclusions vary with the data selection, since results based on the unified and UN database lead to more synergistic SDG 14 and 15 interactions. In contrast, based on WBG database, we find trade-offs (Figure 5B).

Target 12.4 (sound management of chemicals and all wastes throughout their life cycle) has more negative than positive interactions with the other targets in the unified and UN databases. Our results show how various indicator data on electronic waste (Target 12.4) negatively impact or are negatively impacted by other targets. Since neither WBG nor BE-SDSN provides indicator data for Target 12.4, we do not detect those trade-offs in their networks. Due to its potential to impact the environment and human lives, Target 12.4 aims to minimize electronic waste or maximize the amount of e-waste being recycled. However, e-waste is one of the fastest-growing waste streams globally in terms of volume, and only 17.4% of global e-waste is appropriately recycled (Ghimire & Ariya, 2020;

FIGURE 4 Sustainable Development Goal (SDG) target network for (a) synergies and (b) trade-offs based on the unified SDG database. Each node indicates an SDG target, colored according to the respective SDG. The node size reflects the SDG target's eigenvector centrality, emphasizing the structural significance of nodes in the network. Edges indicate synergistic (solid line) or impeding (dashed line) interactions between a pair of targets. The edge's thickness and color reflect the share (s) of synergies and trade-offs (light gray with $30\% < s \leq 60\%$, gray with $60\% < s \leq 90\%$, and dark gray with $s > 90\%$), emphasizing structural significant interactions in the network [Colour figure can be viewed at wileyonlinelibrary.com]



Kumar et al., 2017). Economic growth and urbanization, through access to electricity (Target 7.1) and, therewith, increased access to the internet (Target 17.8), mobile networks (Target 9.c), and mobile

devices (Target 8.10) lead to an exponential generation of electronic equipment (Target 12.4). Many substances in e-waste are toxic (e.g., heavy metals and flame retardants). The improper disposal of

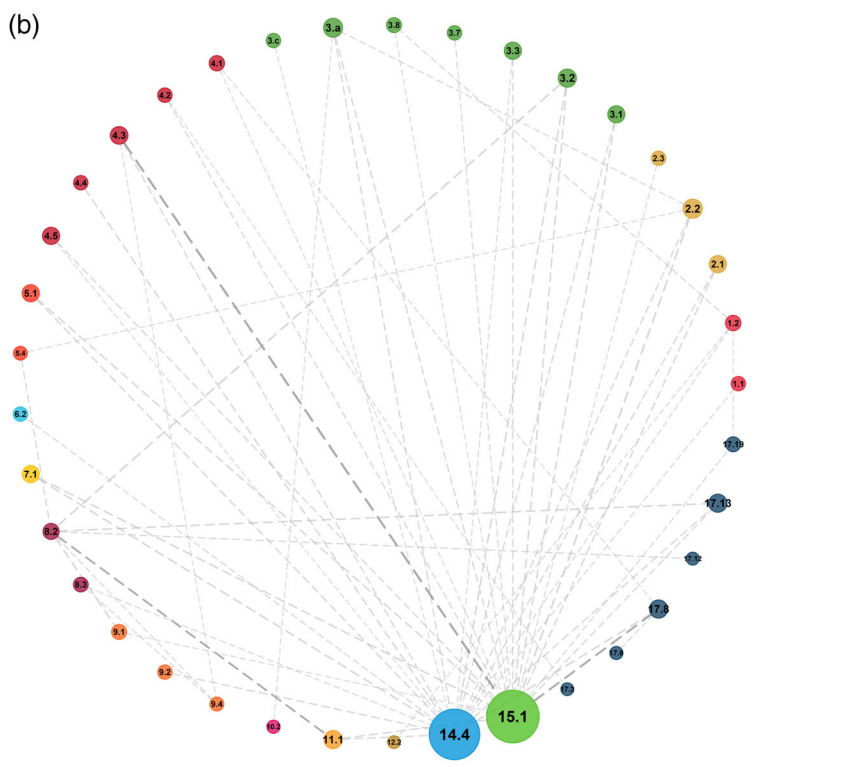
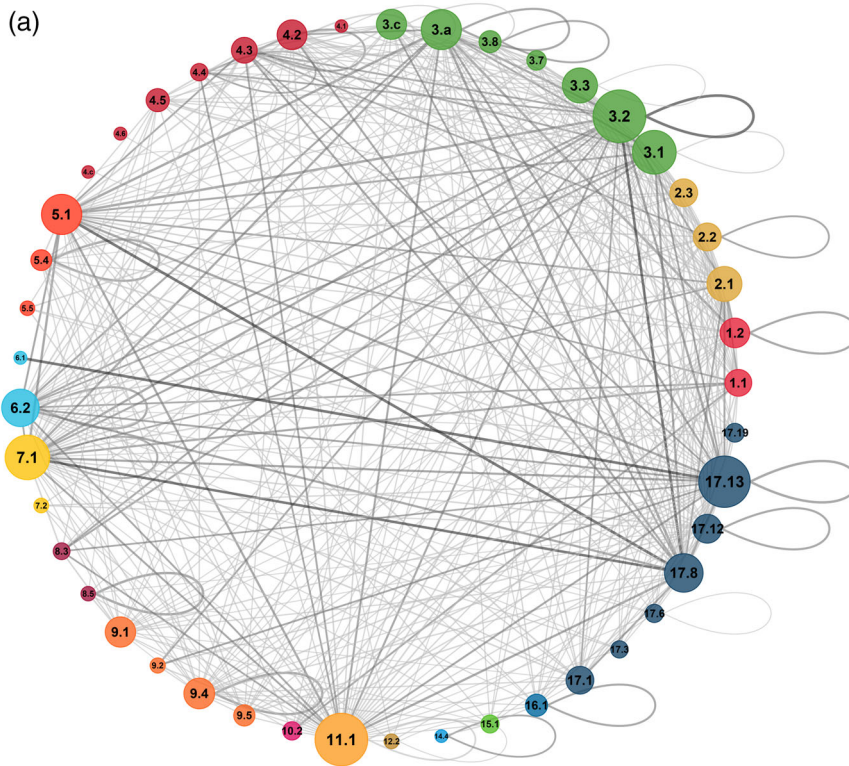


FIGURE 5 Sustainable Development Goal (SDG) target network for (a) synergies and (b) trade-offs based on SDG data provided by the World Bank Group (WBG). Each node indicates an SDG target, colored according to the respective SDG. The node size reflects the SDG target's eigenvector centrality, emphasizing the structural significance of nodes in the network. Edges indicate synergistic (solid line) or impeding (dashed line) interactions between a pair of targets. The edge's thickness and color reflect the share (s) of synergies and trade-offs (light gray with $30% < s \leq 60%$, gray with $60% < s \leq 90%$, and dark gray with $s > 90%$), emphasizing structural significant interactions in the network [Colour figure can be viewed at wileyonlinelibrary.com]

e-waste causes soil and groundwater contamination (SDG 14 and 15), not allowing the farmland to be as productive as possible (SDG 2), and causing illnesses (SDG 3) (Ghimire & Ariya, 2020).

Opportunities associated with e-waste recycling exist, but they are not yet sufficiently exploited to resolve these trade-offs (Ghimire & Ariya, 2020; Kumar et al., 2017).

4 | DISCUSSION

Our study compared SDG interactions based on the UN, WBG, BE-SDSN and unified database to emphasize the role of data in SDG research. We highlight that the data selection for SDGs can change the understanding of SDG interactions even if the same mythologies are applied. Further, this study provided several novel contributions to SDG research by providing new insights on SDG interactions based on different data and scales.

First, our study introduces a framework to build a unified SDG database based on existing ones to fill some of the data gaps and monitor the multifaceted SDGs. Even 5 years after the launch of the 2030 Agenda, a number of the 232 official SDG indicators are still being discussed (Tier II indicator) as well as the monitoring approaches to collect data (Nat Sustain, 2018). Several SDG targets are not covered by data at all or are covered poorly across the existing SDG databases. The reasons for this are manifold. Some of the targets are far ahead of the available statistics, and, in other cases, appropriate statistical concepts from which to generate indicators simply do not exist (MacFeely, 2018). Further, the 2030 Agenda is universal and applies to all countries. However, not all defined targets are even applicable to every country (MacFeely, 2018). Still, data disaggregation needs to improve. In some cases, there is a mismatch between data characteristics (Kraak et al., 2018). Some of the SDGs, especially SDG 13–15, are not well quantified to represent the actual purpose described within their targets and capture multidimensional aspects. In the case of SDG 15, science-based indicators like satellite imagery could be used to measure the coverage of forest, drylands, wetlands and mountain regions (Target 15.1) (MacFeely, 2018). In other cases, the provider of data could use newer methods to measure indicators like remote sensing (Estoque, 2020), citizen participatory approaches (Thinyane & Kirschke, 2019), or big data via machine learning (United Nations, 2021; Vinuesa et al., 2020; Ziesche, 2017).

However, our unified database offers an increased data availability over time and space covering more targets and indicator data being disaggregated. Due to this increased amount of indicator data per target, we could consider more SDG indicator data pairs at all scales. The increased data availability and consistency provide an enhanced representation of the officially adopted global SDG indicator framework. Further, the use of single discrete indicators measuring a multifaceted target also introduces the risk that unmeasured aspects of a target may be ignored and interconnections remain unseen or poorly understood (MacFeely, 2018). Since the unified database offers more indicators per target, the results of SDG interactions are more nuanced and reliable, particularly visible in the network analysis. This robustness becomes evident at the country level through increased data stability throughout the global SDG indicator framework. Researchers and policymakers can use the unified framework and database to monitor SDGs further and explore their interactions.

Second, we detect similarities in SDG interactions among databases globally, at income and regional scales and across networks. Although synergies outweigh trade-offs, our analysis shows that most SDG interactions are not-classified, independent of the data selection.

This finding resembles the one by Warchold et al. (2021) which applied a cross-sectional correlation method using the UN database. However, we still observe more synergies and trade-offs over time than this cross-sectional analysis. Nonetheless, these findings contrast studies that found more synergies and trade-offs than not-classified SDG interactions (Kroll et al., 2021; Pradhan et al., 2017; Weitz et al., 2018). For example, we adopted the correlation method introduced by Pradhan et al. (2017), but did not detect as many synergies and trade-offs. One reason is the choice of a higher threshold for measuring the Spearman's ranked correlation coefficient (ρ). Whereas we measure (ρ) only if at least eight paired observations were available, Pradhan et al. (2017) used three paired observations. Our different choice of intervals to define synergies and trade-offs compared to Pradhan et al. (2017) and the exclusion of dummy variables for the analysis can be other reasons for differences. It also can be due to changes in the SDG data itself over the years after the implementation in 2015. Further, we highlighted that, independent of the data selection, socially orientated SDG interactions are more synergistic, whereas economically orientated SDG interactions are impeding. Pradhan et al. (2017), Kroll et al. (2021), Dawes (2020), and Warchold et al. (2021) also stated that SDG 1–5 show incredibly synergistic interactions with other goals, whereas SDG 8, 12 has the most trade-offs.

Third, we highlighted differences in SDG interactions caused by the data selection while applying the same methodology. Thus, depending on the data selection, research findings may foster actions that might neglect or exacerbate trade-offs. BE-SDSN database shows the highest shares of synergies and trade-offs at all scales, UN database consistently has the lowest shares of synergies and trade-offs, and therefore more not-classified interactions. Moreover, environmentally oriented SDG exhibit contrary results depending on the data selection. The reason is the quite diverse indicator data for SDG 13, 14, and 15 among the databases. For example, the topic of biodiversity (SDG 14, 15) reoccurs in our top 10 differences for synergies and trade-offs at all scales. The classification of those biodiversity-related SDG interactions depends on the data selection, wherefore progress toward their achievements is still lagging. This finding corresponds with the report analyzing the progress of 93 environment-related SDG indicators (United Nations Environment Programme, 2019). The report stated that for 68% of the environment-related SDG indicators there is insufficient data to assess progress and for 9% it is unlikely that the target will be met without increased action (United Nations Environment Programme, 2019). The report hints that indicators without adequate data to assess progress are likely facing a negative trend since they receive less attention in terms of policy interventions and investment in monitoring (United Nations Environment Programme, 2019).

This aspect does not only apply to environmental-related SDGs but also, for example, to the ones covering sustainable consumption and production (SDG 12). SDG 12 is characterized by low data availability (WBG database) or even no data (BE-SDSN database), therewith low funding, and the lowest priority in countries' Voluntary National Reviews and SDG prioritization processes (United Nations Environment Programme, 2019). However, especially our results based on the unified

database reveal that SDG 12 negatively impacts or is negatively impacted by other goals and targets. A similar finding holds for SDG 8. Lafortune et al. (2020) that compares four SDG progress reports in the EU produced by the SDSN, the OECD, Eurostat, and ASviS, confirms our findings on contradicting results for environmental- and sustainable consumption and production-related goals and targets. Despite those contradicting findings on progress towards achieving SDG 8 and 12, the EU valued the potential of progress within those goals by launching a bioeconomy agenda to address potential negative effects (European Commission, 2012, 2018). However, the bioeconomy scenario itself can have positive and negative impacts on other SDGs, wherefore policy regulations need to assure a sustainable bioeconomy strategy which then strongly supports the achievement of all other SDGs (Heimann, 2019; Ronzon & Sanjuán, 2020).

Fourth, our network analysis offers new insights into how the most connected goals and targets vary according to the SDG data selection. The networks based on the unified database provide the most holistic insights into interactions between all goals and 98 target interactions. Our networks based on the WBG database display 12 goals and 50 targets for synergistic and impeding interactions. In comparison, the global SDG target network by Lusseau and Mancini (2019), based on the WBG database, has 71 targets. It should be noted, however, that we only considered targets with shares of synergies or trade-offs that are larger than 30%. This use of thresholds reduces the number of targets in sum but focuses on more connected ones in the network. Independent on the data selection, child mortality is a core target whose mechanisms we need to better understand to progress towards the SDGs. Similar to the findings of Lusseau and Mancini (2019), reducing child mortality (Target 3.2) synergizes with most other SDG targets in the network. However, we observe Target 15.1 and 14.4 to be most negatively connected to the achievement of other targets, which does not become apparent in the results by Lusseau and Mancini (2019). Also, for the goal level, we see similarities in the results for synergies but differences for trade-offs, especially in the case of environmental goals. This finding reinforces the importance to interpret any analytical results of SDG interactions at the same scale as the data used.

The generalization of our findings is constrained by the data availability and our methodological approaches. The assignment of indicator data from the UN, WBG, and BE-SDSN to the unified framework, was in some cases a subjective approach. Since UN and WBG already assigned their indicators to the target level and BE-SDSN at least to the goal level, this subjectivity is somewhat limited. Since we cannot use indicators that are either not well quantified with data (i.e., boolean indicators), some SDG interactions stay unnoticed. Further, our applied correlation analysis does not imply any causality. Since specific SDG interactions occur regardless of the data selection, their causality is still an open field of research. As we only consider bivariate SDG interactions, we can also neither conclude the directions nor multivariate aspects, particularly in the network analysis. A step forward in the network analyses, as done for example by Anderson et al. (2021), are SDG system models reflecting feedback loops and casual relations underlying synergies and trade-offs, where the progressing interaction of some targets might positively or negatively

influence other targets, and goals respectively. We do, however, provide insight into strongly synergistic target connections, which we must take advantage of, which equates to nonlinear dynamics. This means that some SDG will be disproportionately affected by actions to meet other SDGs (Warchold et al., 2021). We attempted to address these drawbacks (causality, direction, and multi-variation) by qualitatively providing literature evidence for the SDG interactions.

5 | CONCLUSION

Building a unified SDG framework and database, we provided an enhanced understanding of data-centric SDG interactions. Our study has outlined some of the major challenges in SDG data application, possible consequences and strategic issues for research emerging for the 2030 Agenda. From a global holistic perspective of evaluating progress towards the achievements of SDGs, it is essential to understand how selecting data might change SDG interactions and consequently the underlying messages. Especially, our findings on income or regional-based SDG interactions show that SDGs should be adapted to specific constraints of countries since the data selection changes even more drastically SDG interactions. The varying data availability, inconsistent data format, and the tension between national and global perspectives make it almost impossible for the data-driven SDG research community to create comparable results by each goal or target. Interferences from data may consciously but also unconsciously be conditioned by their providers as institution pursues their own goals. Therefore, the data and conclusions should be placed in context to minimize the risk of misinterpretation. The data selections can cause aspects of sustainable development to be omitted, oversimplified or overcomplicated, leading to misguided conclusions and investments. Further, the different levels of aggregation can tell different stories – SDG goal versus target level. Therefore, it is necessary to vary the levels of aggregation to confirm tendencies and understand at which point the results diverge or reverse. The above-highlighted obstacles of SDG data should be addressed to prioritize implementation strategies. Mainly, there is a need to develop a unified SDG monitoring framework to implement the 2030 Agenda successfully. At the mean time, efforts are also needed to develop more science-based goals, targets, and indicators for the next generation of global goals for sustainability—the post-2030 Agenda. These goals should ensure societal well-being, prosperity, and public and planetary health and be based on scientific evidence, for example, planetary boundaries, accounting for sustainable governance of global and local commons.

CONFLICT OF INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

ORCID

Anne Warchold  <https://orcid.org/0000-0002-0280-2758>

Prajal Pradhan  <https://orcid.org/0000-0003-0491-5489>

Muhammad Panji Islam Fajar Putra  <https://orcid.org/0000-0002-6729-941X>

Jürgen P. Kropp  <https://orcid.org/0000-0001-7791-3420>

ENDNOTE

ⁱ The unified database does not provide data for following 26 targets: 1.b, 4.7, 7.b, 11.3, 11.4, 11.7, 11.a, 11.c, 12.3, 12.6, 12.7, 12.8, 12.a, 13.3, 13.a, 13.b, 14.1, 14.7, 14.c, 15.7, 15.c, 16.4, 16.7, 17.5, 17.7, 17.14.

REFERENCES

Allen, C., Metternicht, G., & Wiedmann, T. (2021). Priorities for science to support national implementation of the sustainable development goals: A review of progress and gaps. *Sustainable Development*, 29, 635–652. doi:10.1002/sd.2164

Anderson, C. C., Denich, M., Warchold, A., Kropp, J. P., & Pradhan, P. (2021). A systems model of SDG target influence on the 2030 agenda for sustainable development. *Sustainability Science*. <https://doi.org/10.1007/s11625-021-01040-8>

Asadikia, A., Rajabifard, A., & Kalantari, M. (2021). Systematic prioritisation of SDGs: Machine learning approach. *World Development*, 140, 105269. <http://doi.org/10.1016/j.worlddev.2020.105269>

Cardinale, B. J., Duffy, J. E., Gonzalez, A., Hooper, D. U., Perrings, C., Venail, P., Narwani, A., MacE, G. M., Tilman, D., Wardle, D. A., Kinzig, A. P., Daily, G. C., Loreau, M., Grace, J. B., Larigauderie, A., Srivastava, D. S., & Naeem, S. (2012). Biodiversity loss and its impact on humanity. *Nature*, 486, 59–67. <https://doi.org/10.1038/nature11148>

Coenen, J., Glass, L.-M., & Sanderink, L. (2021). Two degrees and the SDGs: A network analysis of the interlinkages between transnational climate actions and the sustainable development goals. *Sustainability Science*. <https://doi.org/10.1007/s11625-021-01007-9>

Dawes Jonathan, H. P. (2020). Are the sustainable development goals self-consistent and mutually achievable? *Sustainable Development*, 28, 101–117. <https://doi.org/10.1002/sd.1975>

Elsevier (2020). The power of data to advance the SDGs. Mapping research for the sustainable development goals. In *Tech. Rep.* Elsevier Foundation, RELX.

Estoque, R. (2020). A review of the sustainability concept and the state of SDG monitoring using remote sensing. *Remote Sensing*, 12, 1770. <https://doi.org/10.3390/rs12111770>

European Commission (2012) *Innovating for sustainable growth* (Technica; Report) European Union (EU). Brussels.

European Commission (2018) *A sustainable bioeconomy for Europe: Strengthening the connection between economy, society and the enviroment* (Technical Report) Europea Union (EU). Brussels.

Fuso Nerini, F., Tomei, J., To, L. S., Bisaga, I., Parikh, P., Black, M., Borrion, A., Spataru, C., Castán Broto, V., Anandarajah, G., Milligan, B., & Mulugetta, Y. (2018). Mapping synergies and trade-offs between energy and the sustainable development goals. *Nature Energy*, 3, 10–15. <https://doi.org/10.1038/s41560-017-0036-5>

Ghimire, H., & Ariya, P. A. (2020). E-wastes: Bridging the knowledge gaps in global production budgets, composition, recycling and Sustainability implications. *Sustainable Chemistry*, 1, 154–182. <http://doi.org/10.3390/suschem1020012>

Hauke, J., & Kossowski, T. (2011). Comparison of values of Pearson's and Spearman's correlation coefficients on the same sets of data. *QUAGEO*, 30, 87–93. <https://doi.org/10.2478/v10117-011-0021-1>

Heimann, T. (2019). Bioeconomy and SDGs: Does the bioeconomy support the achievement of the SDGs? *Earth's Future*, 7, 43–57. <http://doi.org/10.1029/2018ef001014>

Interstate Technology, & Regulatory Council. (2013). *Groundwater statistics and monitoring compliance, statistical tools for the project life*. Cycle. <http://www.itrcweb.org/gsmc-1/>

Jägermeyr, J., Pastor, A., Biemans, H., & Gerten, D. (2017). Reconciling irrigated food production with environmental flows for sustainable development goals implementation. *Nature Communications*, 8, 1–9. <http://doi.org/10.1038/ncomms15900>

Kraak, M., Ricker, B., & Engelhardt, Y. (2018). Challenges of mapping sustainable development goals indicators data. *ISPRS International Journal of Geo-Information*, 7, 482. <http://doi.org/10.3390/ijgi7120482>

Warchold, A., Pradhan, P., & Kropp, J. P. (2021). Variations in sustainable development goal interactions: Population, regional, and income disaggregation *Sustainable Development*, 29, 285. <http://doi.org/10.1002/sd.2145>

Kumar, A., Holuszko, M., & Espinosa, D. C. R. (2017). E-waste: An overview on generation, collection, legislation and recycling practices. *Resources, Conservation and Recycling*, 122, 32–42. <http://doi.org/10.1016/j.resconrec.2017.01.018>

Ladha, J. K., Jat, M. L., Stirling, C. M., Chakraborty, D., Pradhan, P., Krupnik, T. J., Sapkota, T. B., Pathak, H., Rana, D. S., Tesfaye, K., & Gerard, B. (2020). Achieving the sustainable development goals in agriculture: The crucial role of nitrogen in cereal-based systems. *Advances in Agronomy* (vol. 163, pp. 39–116). Elsevier. <https://doi.org/10.1016/bs.agron.2020.05.006>

Lafortune, G., Fuller, G., Schmidt-Traub, G., & Kroll, C. (2020). How is progress towards the sustainable development goals measured? Comparing four approaches for the EU. *Sustainability*, 12, 7675. <http://doi.org/10.3390/su12187675>

Laumann, F., von von Kügelgen, J., & Barahona, M. (2020). Non-linear interlinkages and key objectives amongst the Paris agreement and the sustainable development goals. In *ICLR 2020 Workshop “Tackling Climate Change with Machine Learning”* (Vol. 13. pp. 1–18) arXiv. <http://arxiv.org/abs/2004.09318>

Lewis, T., Synowiec, C., Lagomarsino, G., & Schweitzer, J. (2012). E-health in low- and middle-income countries: Findings from the center for health market innovations. *Bulletin of the World Health Organization*, 90, 332–340. <https://doi.org/10.2471/blt.11.099820>

Lusseau, D., & Mancini, F. (2019). Income-based variation in sustainable development goal interaction networks. *Nature Sustainability*, 2, 242–247. <https://doi.org/10.1038/s41893-019-0231-4>

MacFeely, S. (2018). *The 2030 agenda: An unprecedented statistical challenge* (Technical Report). November, Friedrich-Ebert-Stiftung, New York. <http://library.fes.de/pdf-files/iez/14796.pdf>.

Mainali, B., Luukkanen, J., Silveira, S., & Kaivo-oja, J. (2018). Evaluating synergies and trade-offs among sustainable development goals (SDGs): Explorative analyses of development paths in South Asia and sub-Saharan Africa. *Sustainability*, 10, 815. <http://doi.org/10.3390/su10030815>

Miola, A., & Schiltz, F. (2019). Measuring sustainable development goals performance: How to monitor policy action in the 2030 agenda implementation? *Ecological Economics*, 164, 106373. <https://doi.org/10.1016/j.ecolecon.2019.106373>

Nat Sustain, (2018). Tracking progress on the SDGs. *Nature Sustainability*, 1, 377–377. <http://www.nature.com/articles/s41893-018-0131-z>

Obersteiner, M., Walsh, B., Frank, S., Havlik, P., Cantele, M., Liu, J., Palazzo, A., Herrero, M., Lu, Y., Mosnier, A., Valin, H., Riah, K., Kraxner, F., Fritz, S., & Van Vuuren, D. (2016). Assessing the land resource–food price nexus of the sustainable development goals. *Science Advances*, 2. <http://doi.org/10.1126/sciadv.1501499>

Pham-Truffert, M., Metz, F., Fischer, M., Rueff, H., & Messerli, P. (2020). Interactions among sustainable development goals: Knowledge for identifying multipliers and virtuous cycles. *Sustainable Development*, 28, 1236–1250. <http://doi.org/10.1002/sd.2073>

Pradhan, P. (2019). Antagonists to meeting the 2030 agenda. *Nature Sustainability*, 2, 171–172. <http://doi.org/10.1038/s41893-019-0248-8>

Pradhan, P., Costa, L., Rybski, D., Lucht, W., & Kropp, J. P. (2017). A systematic study of sustainable development goal (SDG) interactions. *Earth's Future*, 5, 1169–1179. <http://doi.org/10.1002/2017ef000632>

- Pradhan, P., Subedi, D. R., Khatriwada, D., Joshi, K. K., Kafle, S., Chhetri, R. P., Dhakal, S., Gautam, A. P., Khatriwada, P. P., Mainaly, J., Onta, S., Pandey, V. P., Parajuly, K., Pokharel, S., Satyal, P., Singh, D. R., Talchabhadel, R., Tha, R., Thapa, B. R., ... Bhuju, D. R. (2021). The covid-19 pandemic not only poses challenges, but also opens opportunities for sustainable transformation. *Earth's Futures*. <http://doi.org/10.1029/2021ef001996>
- Putra, M. P. I. F., Pradhan, P., & Kropp, J. P. (2020). A systematic analysis of water-energy-food security nexus: A south Asian case study. *Science of the Total Environment*, 728, 138451. <http://doi.org/10.1016/j.scitotenv.2020.138451>
- Ritchie, H. & Roser, M. (2019). Access to energy. Our World in Data. <https://ourworldindata.org/energy-access>
- Ronzon, T., & Sanjuán, A. I. (2020). Friends or foes? A compatibility assessment of bioeconomy-related sustainable development goals for European policy coherence. *Journal of Cleaner Production*, 254, 119832. <http://doi.org/10.1016/j.jclepro.2019.119832>
- Sachs, J., Schmidt-Traub, G., Kroll, C., Lafortune, G., Fuller, G., & Woelm, F. (2020). The sustainable development goals and COVID-19. Sustainable development report 2020. In *Tech. Rep.* Cambridge University Press.
- Smarandache, F. (2016). Alternatives to Pearson's and Spearman's correlation coefficients. *SSRN Electronic Journal*. <http://doi.org/10.2139/ssrn.2725499>
- Sustainable Development Solutions Network (2015). Indicators and a monitoring framework for the sustainable development goals: Launching a data revolution for the SDGs. In *Tech. Rep.* Leadership Council of the Sustainable Development Solutions Network. <https://sustainabledevelopment.un.org/content/documents/2013150612-FINAL-SDSN625Indicator-Report1.pdf>
- Thinyane, M. and Kirschke, S. (2019). Data assemblages for enhanced citizen participation in sustainable development. *Global partnership for sustainable development data*. Global Partnership for Sustainable Development Data. <https://www.data4sdgs.org/news/data-assemblages-enhanced-citizen-participation-sustainable-development>
- Tsiafouli, M. A., Thébaud, E., Sgardelis, S. P., de Ruiter, P. C., van der Putten, W. H., Birkhofer, K., Hemerik, L., de Vries, F. T., Bardgett, R. D., Brady, M. V., Bjornlund, L., Jørgensen, H. B., Christensen, S., Hertefeldt, T. D., Hotes, S., Gera Hol, W., Frouz, J., Liiri, M., Mortimer, S. R., ..., Hedlund, K. (2015). Intensive agriculture reduces soil biodiversity across Europe. *Global Change Biology*, 21, 973–985. <https://doi.org/10.1111/gcb.12752>
- UN-Water (2016). Water and sanitation interlinkages across the 2030 agenda for sustainable development. In *Tech. Rep.* UN Water. <https://www.unwater.org/publications/water-sanitation-interlinkages-across-2030-agendasustainable-development/>
- United Nations. Regional groups of Member States. United Nations, Department for General Assembly and Conference Management. <https://www.un.org/dgacm/en/content/regional-groups>
- United Nations (2021). Big Data for Sustainable Development. <https://www.un.org/en/global-issues/big-data-for-sustainable-development>
- United Nations Environment Programme (2019). Measuring progress: Towards achieving the environmental dimension of the SDGs (Technical Report). United Nations Environment Programme (UNEP). Nairobi.
- United Nations Statistics Division (2020). SDG indicators: United Nations global SDG database. URL: <https://unstats.un.org/sdgs/indicators/database/>
- U.S. Environmental Protection Agency (2019). Understanding the science of ocean and coastal acidification. <https://www.epa.gov/ocean-acidification/understanding-science-ocean-and-coastal-acidification>.
- Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., Felländer, A., Langhans, S. D., Tegmark, M., & Fuso Nerini, F. (2020). The role of artificial intelligence in achieving the sustainable development goals. *Nature Communications*, 11. <https://doi.org/10.1038/s41467-019-14108-y>
- Warchold, A., Pradhan, P., & Kropp, J. P. (2021). Variations in sustainable development goal interactions: Population, regional, and income disaggregation. *Sustainable Development*, 29, 285–299. <https://doi.org/10.1002/sd.2145>
- Weitz, N., Carlsen, H., Nilsson, M., & Skånberg, K. (2018). Towards systemic and contextual priority setting for implementing the 2030 agenda. *Sustainability Science*, 13, 531–548. <http://doi.org/10.1007/s11625-017-0470-0>
- World Bank Group (2020a) DataBank: Sustainable development goals (SDGs). [https://databank.worldbank.org/source/sustainable-development-goals-\(sdgs\)](https://databank.worldbank.org/source/sustainable-development-goals-(sdgs))
- World Bank Group (2020b). World Bank country and lending groups. <https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups>.
- Zhang, J., Wang, S., Pradhan, P., Zhao, W., & Fu, B. (2022). Untangling the interactions between the sustainable development goals in China. *Science Bulletin*. <http://doi.org/10.1016/j.scib.2022.01.006>
- Ziesche, S. (2017). Innovative big data approaches for capturing and analyzing data to monitor and achieve the SDGs (Technical Report). ESCAP East and North-East Asia office. ESCAP Statistics Division, Incheon, Republic of Korea. <https://reliefweb.int/sites/reliefweb.int/files/resources/InnovativeBigDataApproachesforCapturingandAnalyzingDatatoMonitorandAchievetheSDGs.pdf>.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Warchold, A., Pradhan, P., Thapa, P., Putra, M. P. I. F., & Kropp, J. P. (2022). Building a unified sustainable development goal database: Why does sustainable development goal data selection matter? *Sustainable Development*, 1–16. <https://doi.org/10.1002/sd.2316>