The Review of International Organizations
https://doi.org/10.1007/s11558-024-09548-1

Tracking earmarked funding to international organizations: Introducing the earmarked funding dataset

Bernhard Reinsberg[1,2] · **Mirko Heinzel**[1,3] · **Christian Siauwijaya**[1]

Accepted: 9 April 2024
© The Author(s) 2024, corrected publication 2024

Abstract

Over the past two decades, donor funding of international organizations (IOs) such as the United Nations, has shifted from core resources toward earmarked contributions. While core resources provide IOs with discretion how to achieve their mandates, earmarked contributions require IOs to focus their assistance on specific themes, sectors, countries, or projects, as determined by the donors. Our understanding of the drivers and effects of earmarked funding has been limited by the lack of systematic data. To address this gap, we introduce the *Earmarked Funding Dataset*. The dataset contains 342,812 earmarked aid activities from nearly 50 donors with over 340 IOs with a mandate in international development from 1990 to 2020. We first explore the data descriptively to cast light on the evolution of the multilateral development system. We then use the data to examine whether and how donors financially reward well-performing IOs. To that end, we combine our data with newly available performance data synthesized from Multilateral Organization Performance Assessment Network (MOPAN) evaluations of 32 IOs in 2009–2020. Using difference-in-difference estimations, we find a significantly positive relationship between outcome performance and earmarked contributions. In contrast, performance does not seem to affect core resources. These patterns suggest that donors reward well-performing multilaterals with more funding, albeit based on development outcomes which are not easily contractible. From a policy perspective, these findings suggest that performance-based approaches to aid allocation—as they are currently practiced—upload the financial risks of organizational reform processes to international organizations.

Responsible Editor: Axel Dreher

Research design and conceptualization: B.R. (50%), M.H. (40%), C.S. (10%); statistical analysis: B.R. (50%), M.H. (10%), C.S. (40%); writing: B.R. (60%), M.H. (20%), C.S. (20%). Author order reflects the significance of their contributions.

Extended author information available on the last page of the article

Published online: 10 May 2024

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

Keywords Earmarked funding · International organizations · Development cooperation · Multilateral aid · Multi-bi aid · Aid allocation · Aid effectiveness · Principal-agent theory · Organizational performance · Multilateral Organization Performance Assessment Network

1 Introduction

Earmarked resources have become an important—if not the dominant—source of funding for international organizations (IOs) such as the UN Development Program, the World Food Program, and UN peacebuilding operations, tasked with promoting development, alleviating humanitarian needs, and promoting peace (Barder et al., 2019; Baumann & Weinlich, 2020; UN 2022). Unlike core resources, earmarked funds provide donors with the ability to restrict the use of funds to specific themes and sectors, regions, countries, or projects (OECD, 2020), which has raised concerns on the capacity of IOs to achieve their mandates and about increased transaction costs of delivering aid (Schmid et al., 2021; Reinsberg, 2023; UN-MPTF, 2019).

Researchers have begun to examine the causes and consequences of earmarked funding. Some argue that donor governments have increasingly earmarked their contributions to IOs to regain control over how IOs spend their development funds (Bayram & Graham, 2022; Reinsberg et al., 2015; Graham, 2023). An additional donor motivation is a desire to address global public bads while overcoming bureaucratic capacity constraints (Bermeo, 2017; Eichenauer & Reinsberg, 2017; Reinsberg et al., 2017). Looking at the effects of this changing resource patterns of IOs, research has found that development projects supported by earmarked funds receive significantly lower success ratings and have smaller development impacts than comparable core-funded projects (Heinzel et al., 2023; Heinzel & Reinsberg, 2024). Echoing these findings, IOs that finance a greater share of their operations from earmarked resources have lower performance ratings (Reinsberg & Siauwijaya, 2023).

Despite our growing knowledge about the causes and effects of earmarked funding, our understanding of its significance in the multilateral system remains incomplete due to a lack of comparable data. While the United Nations has recently upgraded its reporting standards, its earmarked funding statistics cover at most 47 entities over the past decade (UN 2022). The World Bank no longer publicly reports financial flows into its trust funds—the main vehicle through which it receives earmarked funding. Across the multilateral system, a common definition of various types of resources is lacking. This situation is problematic because it distorts our view of how the multilateral system is financed. Given that money translates into power, this lack of knowledge also beclouds our view of who holds the strings in this system.

To address these knowledge gaps, we introduce the *Earmarked Funding Dataset*—the most comprehensive data source of its kind. At its lowest level of analysis, it identifies 342,812 earmarked aid activities from nearly 50 donors with over 340 IOs in development from 1990 to 2020. The data has three distinct advantages over alternative datasets. First, the data extends the time-series coverage on donor-funded activities in international organizations in the Creditor Reporting System (CRS)—the workhorse dataset for development researchers that is maintained by the Organiza-

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

tion for Economic Cooperation and Development (OECD). Second, it offers a systematic measure of the stringency of earmarking. Hence, the data allows researchers to examine differences in how donors restrict the discretion of IOs (Baumann et al., 2019; Baumann & Weinlich, 2020; Reinsberg, 2016). The data allow us to capture the stringency of earmarking along three dimensions—aid geography, aid themes, and institutional modalities—and distinguish three stringency levels within each dimension. Third, the data offers readily useable datasets at different levels of aggregation that make related analyses as user-friendly as possible.

The data allow scholars to probe the allocation of earmarked funding at different levels of analyses. First, scholars interested in comparative IO research can use aggregated IO-year data to probe how funding arrangements affect key IO-level outcomes, like institutional design features or policy outputs (Dellmuth et al., 2022; Hooghe & Marks, 2015; Lall, 2023; Lundgren et al., 2023; Zürn et al., 2021). Second, the data could prove useful for those interested in the interaction between domestic politics and the delegation to IOs. By aggregating our data to the donor-year level, they can probe questions on how key transformations in domestic politics affect how donors allocate funding to IOs (Bayram & Graham, 2022; Brutger & Clark, 2022; De Vries et al., 2021; Milner & Tingley, 2013). Third, researchers focusing on differences in aid allocation and effectiveness could utilize our data at the donor-recipient-year level to better understand how different types of assistance are allocated and probe how funding modalities shape aid effectiveness (Dietrich, 2021; Fleck & Kilby, 2006; Jones & Tarp, 2016).

To demonstrate the usefulness of the data, we revisit a question that has been hotly debated in the development community but for which corresponding analysis has been plagued by lack of detailed data. Specifically, we examine whether and how donors respond to performance information about international organizations. Do they increase their financial contributions to international organizations following good performance? We examine this question in the context of performance evaluations from the Multilateral Organization Performance Assessment Network (MOPAN)—a network of 20 donors represented in the OECD Donor Assistance Committee (DAC) that regularly evaluates international organizations, purportedly with a view to encourage reforms and inform funding decisions by donors (Lall, 2021).

Our empirical analyses demonstrate how differentiating types of funding leads to different conclusions on this important empirical question. First, when analyzing the impact of performance evaluations on donor funding in the aggregate, scholars would conclude that donors do not care about external evaluations of IO performance. However, these findings are misleading. The vast majority of core funding is allocated through assessed contributions and regular replenishment negotiations that do not provide the flexibility for donors to react to new information. Hence, it is central to differentiate between funding arrangements. Second, when doing so, we find that donors allocate more earmarked funding after an organization has received a positive performance evaluation. In contrast, we find no evidence that donors alter their *core* contributions in response to these evaluations. Finally, we demonstrate how our highly disaggregated data can aid causal inference through an additional analysis at the donor-IO-year level. These models demonstrate our results are driven

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

by donor countries that are MOPAN members—those member states who have privileged access to and interest in external evaluations.

Our contributions are threefold. First, our data allow for a more accurate description of funding flows in the multilateral system. To our knowledge, we provide the most accurate depiction of earmarked flows in the multilateral system. We also explicitly account for the earmarked outflows of global multilateral funds to operational IOs. Official statistics do not report these flows and thereby underestimate the true amount of earmarked funding for operational IOs. Second, our data allow for empirical tests of when and how donor principals control international development agencies through the 'power of the purse' (Hawkins et al., 2006; Patz & Goetz, 2019; Schneider & Tobin, 2016; Thorvaldsdottir and Patz, 2019; Weaver, 2007). While much of this literature has considered that donors—if not entirely (de-)selecting organizations from their portfolio—adjust total funding to multilaterals based on their perceived efficiency, we highlight a more subtle policy option whereby donors continue to engage but with different funding types and levels of strings attached. We present a measure of earmarking stringency that is systematic, valid, reliable, and replicable. Finally, we contribute new evidence on the relationship between organizational performance and donor funding. We find that—unlike core funding—earmarked funding reacts systematically to performance information, thereby underpinning the notion that donors indeed reward performance.

2 A new dataset on earmarked funding

IOs are the backbone of the multilateral system. They help the international community address global development challenges such as poverty and inequality, fragility and conflict, and climate change. To fulfill their mandate, IOs need resources. In recent decades, the resourcing of IOs has fundamentally changed. During most of the Cold War, IOs traditionally relied on *core resources*—through assessed contributions and unrestricted voluntary contributions (Graham, 2023). After the Cold War, core funding to IOs stagnated, giving rise to earmarked contributions. These earmarked resources have become an important—if not the dominant—source of funding for IOs such as the UN Development Program, the World Food Program, and UN peacebuilding missions (Passmore et al., 2023; Reinsberg et al., 2015; United Nations, 2022). Unlike core resources, earmarked funds provide donors with the opportunity to restrict the use of funds to specific themes and sectors, regions, countries, or projects (OECD, 2020).

Principal–agent models provide a useful starting point for understanding the drivers of earmarked funding. This literature emphasizes how donors delegate aid programs to international bureaucracies to benefit from their global capacity, technical expertise, and policy credibility (Abbott & Snidal, 1998; Hawkins et al., 2006; Rodrik, 1995). However, agency slack may undermine these delegation benefits. Two common forms of agency slack discussed in the literature are *agency slippage*—the pursuit of policy objectives that are not intended by donors—and *agency shirking*—where agents do not exert effort on behalf of their principals (Conceição-Heldt, 2017; Michaelowa et al., 2018; Nielson & Tierney, 2003).

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

To mitigate agency slack, principal–agent theorists discuss several control mechanisms (Hawkins et al., 2006). Donors seek to prevent agency slack through *ex-ante* control mechanisms, such as designing contracts that prescribe narrower mandates. For example, the creation of pass-through IOs—such as the Global Fund, the Global Partnership on Education, and the Green Climate Fund—can be seen as an instance of ex-ante control, whereby donors establish multilaterals that support vertical interventions targeting specific sectoral challenges. In addition, *ex-post* controls seek to detect errant agent behavior. Ex-post mechanisms include monitoring and evaluation, agency sanctions, and under-staffing (Dijkstra, 2014; Eckhard et al., 2023; Reykers & Adriaensen, 2022). As a last resort, donors may also discontinue funding—if not leaving the organization altogether (Von Borzyskowski & Vabulas, 2019).

A common assumption is that donors will cut their funding when agencies are slacking. This assumption overlooks, however, that funding cuts can prove difficult in practice. For example, where organizations fulfill vital roles and are the only hubs of expertise to address a given policy problem, donors may have incentives to continue financial support despite the problems (Lall, 2021). Funding cuts may also be difficult as they would imply a breach of commitment to share funding burdens with other donor countries. An alternative to discontinuing funding is to alter the composition of funding toward funding types that provide donors with greater control over organizational processes.

Synthesizing a burgeoning literature, we would expect that donors use earmarked funding to mitigate the risk of agency slack (Eichenauer & Reinsberg, 2017; Graham, 2017; Sridhar & Woods, 2013). Earmarked funding allows donors to support specific development purposes (OECD, 2012). It also allows donors to specify how IOs must manage their contributions, for example by requiring that funds be managed in separate accounts with tailored financial reporting and more frequent results reporting against donor-defined results frameworks (Reinsberg, 2016). Arguably, these checks limit the discretion for agencies to do as they please, in turn enhancing opportunities for donors to exercise control over agency activities.

Empirical tests of these predictions from principal-agent models require systematic data on earmarked funding for a wide range of IOs over long time periods. Such data is currently lacking. For a long time, the United Nations did not have a common typology of organizational resources, which made budget comparisons between different UN entities difficult (Patz & Goetz, 2019). While the United Nations has recently harmonized its reporting standards, its earmarked funding statistics cover at most 47 entities over the past decade (United Nations, 2022). The UN Department for Economic and Social Affairs holds funding data of 22 UN Development System entities from 2003 to 2020. Across the multilateral system, a common definition of various types of resources is lacking. This is problematic because it distorts our view of how the multilateral system is financed and undermines our ability to test pertinent theories of donor control in IOs.

The Creditor Reporting System (CRS)—the official data source maintained by the OECD/DAC Secretariat—records activity-level information on earmarked aid that is continuously updated (OECD, 2021). However, the CRS has three key limitations. First, it underestimates the number of earmarked aid activities because information on implementing channels is unavailable before the mid-2000s, which makes it dif-

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

ficult to understand longer-term trends in IO resourcing. Second, the CRS does not assess the stringency of earmarking, which prevents nuanced understanding of the differences in funding flows. Third, because it draws on donor-reported information, the CRS contains occasional errors and coding inconsistencies, especially in earlier years (Eichenauer & Reinsberg, 2017).

Researchers began to remedy these shortcomings, leading to the release of the 'multi-bi aid dataset' (Eichenauer & Reinsberg, 2017). This dataset identified 109,271 under-reported earmarked aid activities from over 30 donors between 1990 and 2012 but has not been updated since. Hence, the multi-bi aid data has a fairly limited coverage in terms of donors and for more recent years. The Earmarked Funding Dataset expands on the multi-bi-aid dataset in three ways. First, it updates the yearly coverage to 1990–2020. Second, it expands the list of covered IOs by 47 IOs. Third, it provides new measures of earmarking stringency, including a latent index of earmarking stringency developed in this article.

2.1 Overview of the earmarked funding dataset

The *Earmarked Funding Dataverse* extends earlier data collections on earmarked aid in critical ways. At the most disaggregated level, the dataset includes 342,812 earmarked aid activities from nearly 50 donors with over 340 development IOs from 1990 to 2020. Its unique feature is to augment the CRS source data with several variables that allow researchers to systematically assess the stringency of earmarking across a thematic dimension, a geographic dimension, and an institutional dimension. In addition, the dataset improves upon CRS source data by identifying additional earmarked activities and correcting coding errors and reporting inconsistencies, especially in the earlier years when reporting quality was poorer.

The Earmarked Funding Dataverse includes four components, each incorporating unique features that improve upon existing data collection efforts in terms of validity, reliability, and versatility:

- Component 1 is a cross-sectional list of all IOs with a mandate in development to which donors have provided earmarked funding in the past 30 years. To identify development IOs, we follow the OECD/DAC definition and consider IOs that are eligible to receive Official Development Assistance (ODA). The unique feature of this component is to provide the (to date) most extensive list of earmarked funding channels that also includes prominent institutional sub-accounts. Overall, we identify 730 institutional sub-accounts—typically agency trust funds.[1]
- Component 2 is an activity-level dataset on earmarked funding flows to IOs. It allows for the most fine-grained analysis of earmarked funding patterns. Based on the CRS dataset, it covers 342,812 earmarked aid activities from nearly 50 donors with over 340 development IOs from 1990 to 2020. Its unique feature is to augment the CRS source data with several variables that allow researchers to

[1] In the context of IOs, we understand trust funds as accounts separately managed from core resources that allow IOs to receive donor contributions for special purposes as set out in the funding agreements. They are also called 'agency trust funds' to delineate them from independently governed global trust funds.

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

systematically assess the stringency of earmarking across a thematic dimension, a geographic dimension, and an institutional dimension. In addition, the dataset improves upon CRS source data by identifying additional earmarked activities and correcting coding errors and reporting inconsistencies, especially in the earlier years when reporting quality was poorer.

- Component 3 records the aggregated contributions given by donors in each year. It will be most useful for researchers examining the aid allocations of donor countries. Its main innovation is to provide readily usable aggregates of earmarked aid. These aggregates include separately the earmarked contributions that donor governments provide directly to implementing IOs, as well as the repatriated outflows of pass-through IOs in proportion to donor funding shares in these multilaterals. The component also incorporates allocations of bilateral aid and multilateral aid after any adjustments for multi-bi aid that are necessary in order to avoid double-counting of aid.
- Component 4 is at the IO-year level. It records the total earmarked flows to 315 development IOs from 1990 to 2020. A unique feature of this component is that it allows researchers to assess the stringency of earmarking by tracking how much donors contributed across different earmarking dimensions. The component also includes CRS source data on core funding for a range of IOs. This component will be particularly useful for comparative IO researchers.

We describe the components in greater detail below and provide basic information about how we collected the data. The supplemental appendix provides more detailed information on our coding protocol and data transformations.[2] In the following, we illustrate some key descriptive findings from the various components of the Earmarked Funding Dataset. For example, we will gain new insights into the institutional structures within IOs through which states cooperate for global development endeavors. We will also present estimates of how much earmarked funding donors provide. Finally, we will examine how much donors earmark their contributions to IOs.

2.2 The institutional make-up of the multilateral development system

Component 1 of the data provides an overview of the IOs that can receive ODA and includes the major institutional sub-accounts through which these IOs receive earmarked contributions. While the list of ODA-relevant IOs comes from the DAC list, we enhance the data by coding basic characteristics such as year of establishment, geographical mandate, sectoral mandate, and whether the entity is an operational IO or a pass-through IO. We also add the COW IGO code for courtesy, allowing researchers to merge additional variables from the COW IGO dataset (Pevehouse et al., 2021). In addition, we identify institutional sub-accounts for the major IOs, including the World Bank, the regional development banks, and the largest UN entities. We identify these sub-accounts from IO websites as well as an analysis of

[2] Importantly, our coding decisions follow a systematic script that could be applied to future vintages of the CRS and which would extend the data in time.

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

donor-reported project descriptions in the activity-level CRS data. We code start year, geographical mandate, and sectoral mandate of these sub-accounts, using the harmonized DAC typology for the latter two variables to make comparisons with the parent organizations as straightforward as possible.

Our data provides insights into the institutional make-up of the multilateral development system. Figure 1 shows the number of new IOs established over the course of successive five-year periods following the Second World War. The figure illustrates that the number of additional IOs entering the system has been fluctuating, with establishments peaking in the early-1970s and the post-Cold War period. New development IOs have increasingly been created in the shape of pass-through IOs since the end of the Cold War. In absolute terms, their heyday has been around the Millenium turn. In relative terms, their prominence has continuously increased, given that the birth rates of operational IOs have drastically declined.

Figure 2 shows the number of institutional sub-accounts in our data, available for the most important development IOs. We distinguish two types of sub-accounts. Sub-institutions are specific departments within larger organizations, which also include convention secretariats. Trust funds are financial sub-accounts that allow organizations to keep funds separately from core resources for specific purposes and to pool funds from multiple donors. Two trends stand out. First, the total number of sub-accounts across all development IOs has been increasing, particularly after the Cold War when their growth further accelerated. Second, sub-institutions accounted for much of this growth until the mid-1990s, but trust funds became the dominant form of sub-account after the mid-1990s—in line with findings already established in the literature (Graham, 2023; Reinsberg, 2017; Reinsberg et al., 2015)

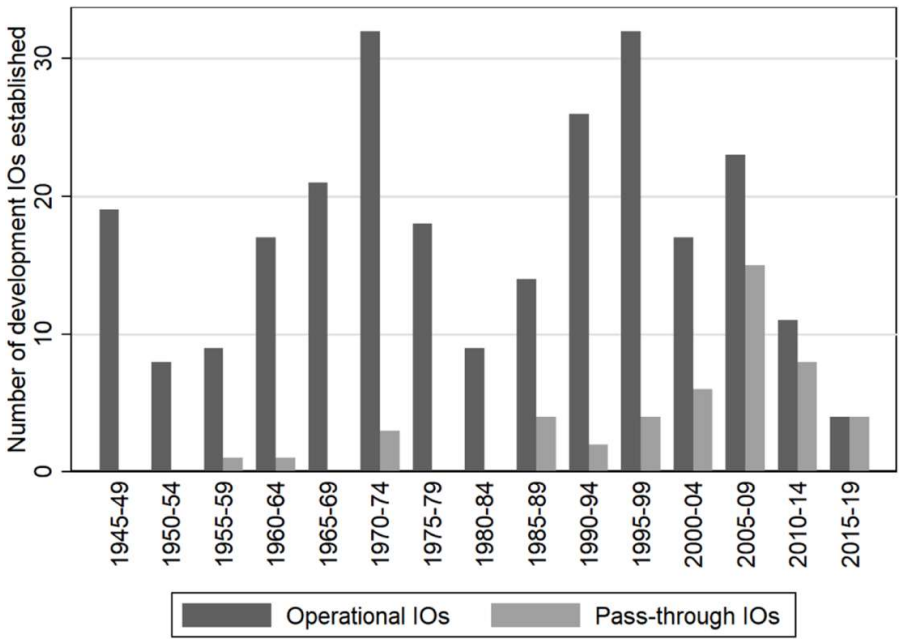


Fig. 1 Number of development IOs over time. *Notes: Operational IOs* are development IOs with capacities for implementation. *Pass-through IOs* are development IOs whose only purpose is to mobilize funding and which rely on operational IOs for implementation

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

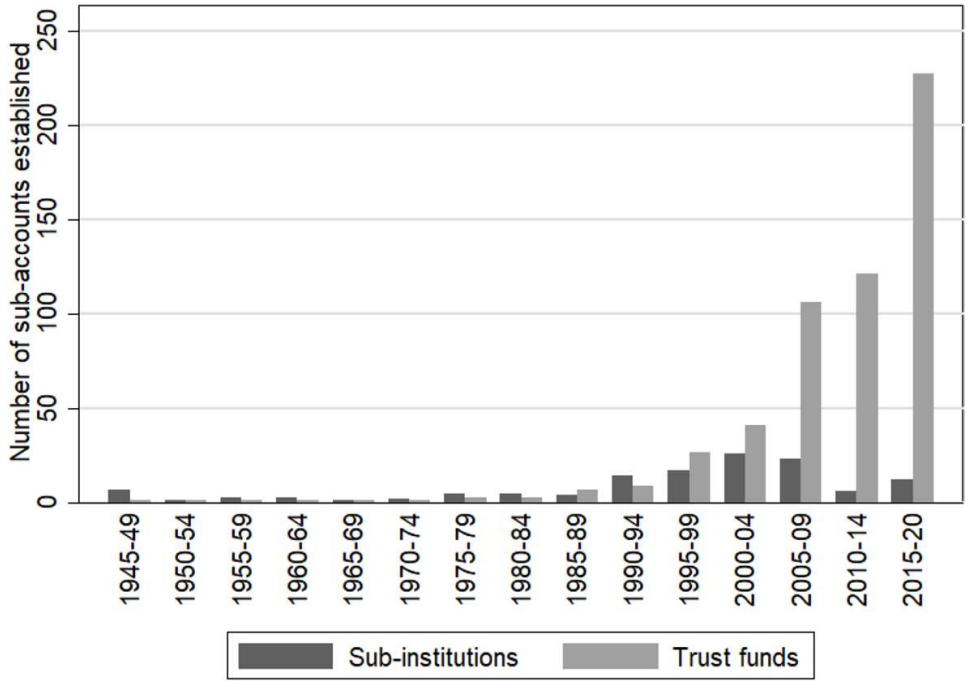


Fig. 2 Number of institutional sub-accounts over time. *Notes:* In practice, *sub-institutions* may receive contributions through *trust funds*. In these few cases, we code both dummies as one

2.3 Global trends in earmarked funding

Component 2 provides the basis for estimating the total flows of earmarked funding in the multilateral system. We identify earmarked activities based on several pieces of information in the CRS database. Earmarked aid activities have an international development organization in the implementing channel.[3] Where the channel information is missing, we turn to project descriptions, using manual coding and a computer-assisted pattern matching algorithm to identify the channel institution. This procedure helps us identify underreported earmarked aid activities. Despite our best efforts, our data likely underestimates the amount of earmarked funding, especially in earlier years. This is because of missing data in the channel institution and underreporting of aid in the CRS data.[4] Our recommendation therefore is to interpret absolute amounts in earlier years with care. The use of funding shares, derived from CRS data, could mitigate the general underreporting problem as we do not expect such underreporting to be systematic across different types of aid.[5] We also recommend that authors use year-fixed effects in their estimations to account for differences in data availability across years.

Component 3 aggregates the data to show the annual earmarked flows from individual DAC donors. Figure 3 displays the evolution of earmarked funding, separately

[3] Where the implementing channel is a generic one for a multilateral institution, we still count this activity toward earmarked aid but cannot code information on earmarking stringency.

[4] Following communication with the OECD/DAC Secretariat, we confirmed that donors began aligning their CRS reporting with their aggregate reporting only in 2012.

[5] A remaining source of underestimation is that we can identify earmarked activities only from project descriptions in 1990–2005, as reporting on channel institutions began only in 2006.

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

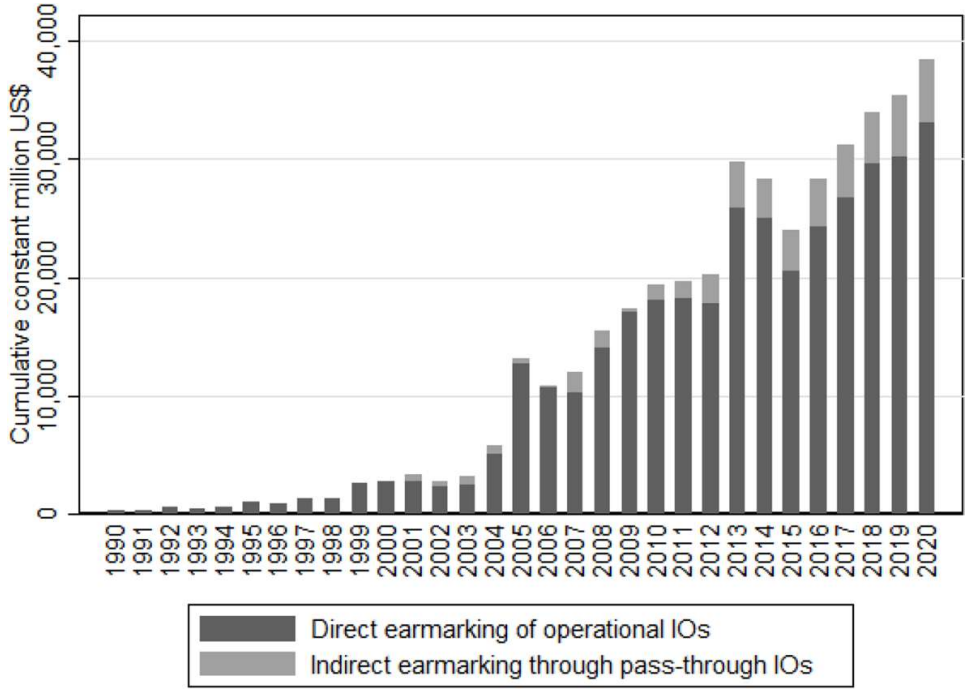


Fig. 3 Evolution of earmarked aid channels. *Notes:* Direct earmarking refers to contributions from bilateral donors to operational IOs. Indirect earmarking refers to earmarked outflows from pass-through IOs as reported in the CRS

for the earmarked aid that DAC donors directly provide to implementing organizations (black bars) and the earmarked aid that can be attributed to them through their shares in pass-through IOs (grey bars). The figure illustrates that earmarked aid has continued to increase over the past decade, albeit more linearly rather than exponentially. In 2020, DAC donors provided $33.0 billion in directly earmarked aid to operational IOs (thereof $27.1 billion by DAC governments and $4.4 billion from the European Commission). This figure is close to the $31.5 billion reported in the DAC1 table (OECD, 2021).

Our data also show that an increasing portion of earmarked DAC aid comes 'through the backdoor', from major pass-through IOs. In 2020, the earmarked outflows of these institutions that were made possible by DAC replenishments reached $5.4 billion. In this figure, we underestimate the amount of earmarked funding from pass-through IOs because we disregard their earmarked outflows that could be attributed to non-DAC donors. Moreover, the true amount of earmarked aid is higher because we do not have earmarked outflows from pass-through IOs that do not report outflows or do not publish data on shareholdings, like the Global Partnership on Education. Our pass-through IOs also do not include inter-agency funds of the UN system, like the Peacebuilding Fund. Hence, our estimates of indirect earmarking through pass-through IOs should be interpreted as a lower bound. More accurate estimates would require that more pass-through IOs would report their earmarked outflows to operational IOs.[6]

[6] There are two options to increase these estimates. One—beyond the scope of this article—is to manually collect DAC contribution data to non-reporting pass-through IOs and count those as earmarked outflows to operational IOs. This would be valid only if the pass-through IOs indeed worked through operational IOs.

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

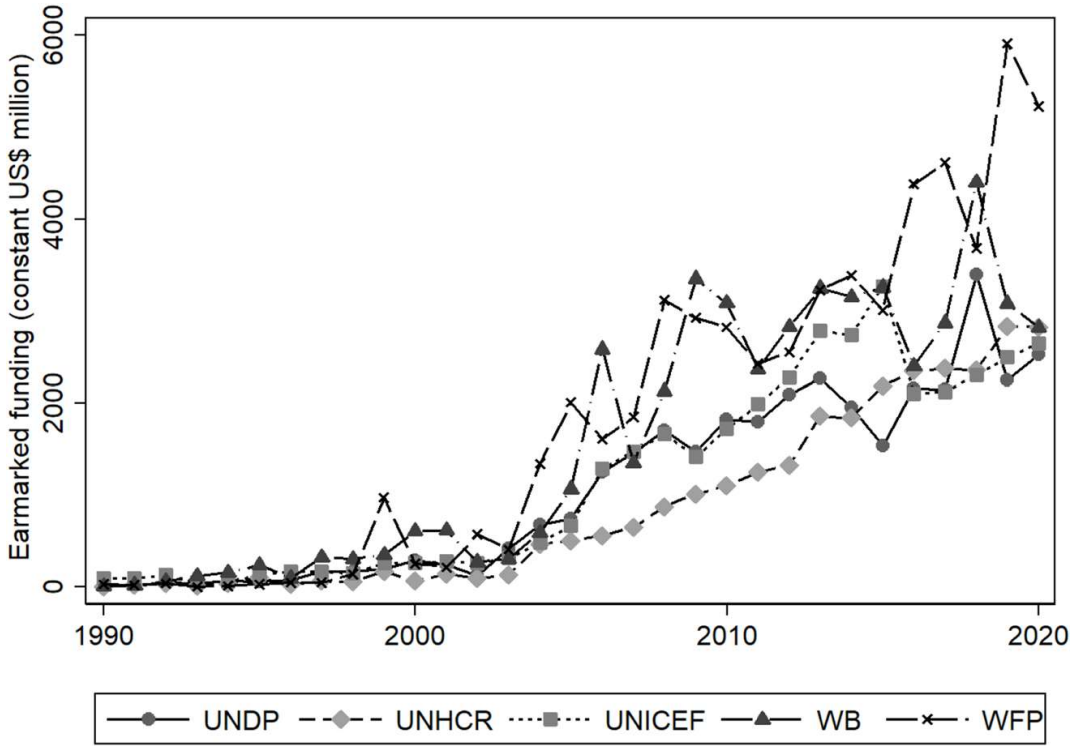


Fig. 4 Top-5 IOs in terms of earmarked funding

Component 4 aggregates the data from the perspective of operational IOs. This allows us to track the earmarked resources to specific IOs. Figure 4 shows the evolution of earmarked funding to the five IOs that received the most earmarked funding in 1990–2020—UN Development Program (UNDP), UN High Commissioner for Refugees (UNHCR), UN Children's Fund (UNICEF), World Bank (excluding IFC), and World Food Program (WFP). Two findings stand out. First, all major IOs increased their earmarked budgets, albeit at different paces and with different trends. While all UN agencies show a clear growth trend, the World Bank seems to have reduced its intake in earmarked funding. WFP had the largest short-term increase in the past few years, which is explained by exceptional crises such as the Covid-19 pandemic and the global food price crisis. Second, the large year-to-year fluctuations in earmarked funding commitments suggest that earmarked contributions are relatively volatile. This high level of volatility is surprising given that total contributions would be expected to look more stable for such globally-operating multi-sector organizations.

2.4 Global trends in the stringency of earmarking

A unique feature of the Earmarked Funding Dataset is to augment the CRS source data with several variables that allow researchers to systematically assess the stringency of earmarking across three dimensions. In each dimension, an activity can

Even then we would not know which operational IOs received these earmarked outflows. Another option is to count CRS-reported donor contributions to non-reporting pass-through IOs as indirect earmarking. Our estimate of indirect earmarking would then be $7.6 billion (compared to $5.4 billion) and direct earmarking would be $30.9 billion (compared to $33.0 billion) (section A.5.1).

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

be strictly earmarked, softly earmarked, or not earmarked at all. First, *geographic earmarking* limits the geographic scope of an activity within the IO mandate. For example, for a global IO, an activity in a specific world region is softly earmarked, whereas a country-specific activity is strictly earmarked. No earmarking exists when the geographical scope of a contribution exactly matches the IO mandate. Second, through *thematic earmarking*, donors restrict the (sub-)sector focus of an activity. Strict earmarking involves specific sub-sectoral interventions, whereas soft earmarking involves broad themes. No earmarking exists when the contribution can be used to support any activity within the IO mandate. Third, *institutional earmarking* restricts IO autonomy in other ways. Soft earmarking specifies funds to benefit specific departments and specific positions within the IO bureaucracy. Strict earmarking involves the secondment of bilateral donor staff into the IO bureaucracy. We provide real-world examples in the codebook. In earlier sample years, we rely on project descriptions to hand-code earmarkers. In later years, we use additional CRS variables to triangulate the earmarking variables. For example, we can identify country-earmarked activities easily by comparing the recipient location with the geographical mandate that we coded in the first component. For thematic earmarking, the CRS includes an aid-type variable that distinguishes between basket funding and project-type interventions. For institutional earmarking, we also use computer-assisted keyword searches for organizational sub-units and specific modalities of institutional collaboration.

To describe the stringency of earmarking in development IOs, we count the number of earmarks across all three dimensions. We assign one point to a soft earmarker and two points to a strict earmarker in any dimension. This yields a theoretical range of a naïve additive earmarking index between 1 (lowest earmarking) and 7 (highest earmarking). Figure 5 presents a breakdown of the total earmarked funding in the

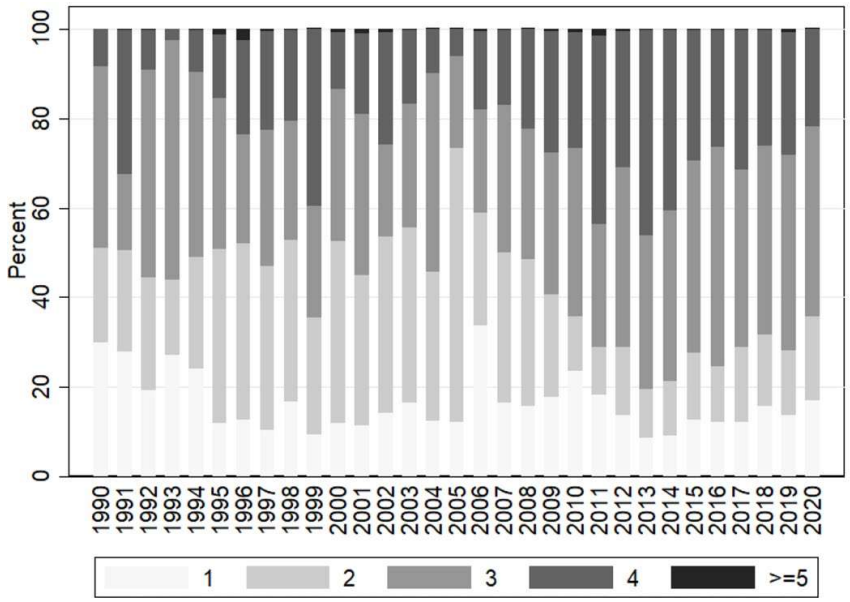


Fig. 5 The distribution of earmarked funding at different levels of earmarking stringency. *Notes:* Earmarking stringency ranges from 1 (soft earmarking in only one dimension) to 7 (strict earmarking in all three dimensions and for an organizational sub-unit). The highest three levels are infrequent and have thus been combined in one bin

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

entire multilateral system across different levels of earmarking stringency. The data show the most stringent kinds of earmarking with at least five earmarkers play virtually no role. However, many activities still have four earmarks, like country-specific project-type interventions. Furthermore, the number of earmarkers with the most cumulative funding differs over distinct periods. For example, in 2009-13, nearly the same amounts of earmarked funding went to activities with three earmarks and activities with four earmarks following a sub-period of progressively stricter earmarking after a phase in which donors provided a lot of multilateral debt relief. In 2014-20, the amount of funding with four earmarks has declined relative to the amount with three earmarks, roughly returning to the proportions seen in 1990-99. Third, while the category with three earmarks has increased in recent years, activities with just one earmark have consistently expanded, too, since their relative low in 2013. This could be interpreted as evidence that efforts to pool earmarked funds in larger thematic facilities have come to bear fruit, although the pace of growth is slow.

2.5 What have we learned?

Our descriptive analysis of the Earmarked Funding Dataset offers new insights into the financial composition of the multilateral development system. We have shown that earmarked funding has become a significant source of funding and that an increasing portion of such earmarked aid comes from pass-through IOs. We also found evidence of increasing earmarking stringency, with a parallel trend toward less-earmarked contributions at smaller scale. These findings extend the insights that can be gained from alternative sources of funding data, like the OECD/DAC CRS (OECD, 2021) and the UN Data Cube (UN 2022).

Compared to the CRS, our data indicates higher amounts of earmarked funding. These differences appear because we remedy coding errors and count the resources of pass-through IOs as earmarked outflows—at the point of entry to operational IOs—not as unearmarked multilateral inflows from bilateral donor governments. For analyses on IO performance, this methodological choice is preferrable. For analyses of donor aid allocation decisions, a focus on multilateral inflows to pass-through IOs may be adequate. Our datasets are sufficiently flexible for either choice, which is an advantage over the CRS data.

Compared to the UN Data Cube, our data will likely underestimate the amount of earmarked funding. This is because we restrict the set of reporting donors to DAC members, which include governments and the European Commission. However, UN agencies draw on a wider support base, including private donors, recipient countries, and own revenue, for instance from patent licensing in the case of the World Intellectual Property Organization. Hence, scholars who are only interested in the diverse nature of UN resources should use the UN data source.

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

3 An empirical application

In this section, we provide an empirical application of the earmarked funding data to illustrate its usefulness for applied research on IOs. We examine whether and how donors reward IOs for good performance. This question has been debated in the development community, but lack of data made related analysis difficult. After reviewing the theoretical debate and the existing evidence, we describe our data and research design and present the results.

3.1 Theoretical expectations

The rise of performance rankings has been a striking phenomenon in global governance (Broome et al., 2018; Honig & Weaver, 2019; Kelley & Simmons, 2015; Palagashvili & Williamson, 2021). Global governance actors seek to exert soft pressure by producing metrics that make it easier to compare across countries or organizations. The targets of these soft pressures are nudged into changing their behavior or policy practice as rankings make performance deficits explicit. Examples for governance through ratings abound. The Doing Business report of the World Bank has infamously served to make developing countries into model students of neoliberal policymaking (Broome et al., 2018). Similarly, the Aid Transparency Index has increased pressures on foreign aid donors to be more transparent (Honig & Weaver, 2019). Finally, both bilateral donors and multilateral bodies—notably the Multilateral Organization Performance Assessment Network (MOPAN)—have begun assessing the organizational performance of IOs, purportedly with a view to encourage reforms and inform funding decisions among performance-oriented donors (Lall, 2021).

We focus on the case of MOPAN performance ratings to study whether donors reward IOs for good performance. Anecdotal evidence suggests that at least some donors adjust their multilateral contributions in response to performance information. For example, following its 2011 Multilateral Aid Reviews, the UK has phased out core funding to purportedly ineffective agencies like the International Labour Organization (ILO), the UN Human Settlements Programme (UN-Habitat), the UN International Strategy for Disaster Reduction (UNISDR), and the UN Industrial Development Organization (UNIDO). Beyond these high-profile cases, Lall (2021) has examined the link between performance assessments and financial contributions. He showed that donors reward good performance with greater total budgetary contributions, but only if agencies provide services that make them relatively exchangeable with competitors and if agencies do not have strong alliances with non-state actors that would allow these agencies to undercut responsiveness.

We re-examine the link between performance assessments and donor contributions. Our Earmarked Funding Dataset affords us with the possibility for more fine-grained analysis to ascertain not only whether but *how* donors reward performance. Theoretically, we might expect donors to earmark more after a poor evaluation to correct perceived deficits. Here we assume that donors still want to engage with the organization. It is also possible that donors may substantially cut their funding (Hirschmann, 2021; Hüfner, 2017; Patz & Goetz, 2019). As a last resort, donors might even disengage from the organization (Von Borzyskowski & Vabulas, 2019).

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

These alternative responses are beyond the scope of our empirical tests and we focus only on responses that involve a change in financial contributions.

If donors observe a positive evaluation, they may provide additional funding. Paradoxically, such funding increases may come as earmarked resources. Good performance reassures donors that the agency is run effectively, which could prompt them to release short-term funding that has already been pre-disposed for specific purposes. As corroborated in background interviews with donors, this prediction is plausible because core funding is locked in by inter-state agreements and cannot be increased at the discretion of an individual donor. For example, a FCDO interviewee said that earmarked funding was an attempt to "save the multilateral system", using funding mechanisms as a tool to "reward good performers while at the same time responding to domestic pressures for greater accountability over multilateral contributions." In sum, our primary expectation therefore is that donors reward good performance with a short-term increase in earmarked funding. A secondary expectation would be that good performance may also increase core contributions in the long run.

3.2 Data and methods

3.2.1 Datasets

To examine the relationship between organizational performance and financial contributions from OECD/DAC donors, we combine the *Earmarked Funding Dataset* with the *MOPAN Performance Dataset*. From the Earmarked Funding Dataset, we draw on the IO-year level component, which records the annual total earmarked flows to 315 IOs from 1990 to 2020 along with a measure of the average stringency of earmarked activities. This component already includes the core contributions from OECD/DAC donors. They are available from the OECD/DAC statistics on the use of the multilateral system for most IOs from 2011 to 2020.

The *MOPAN Performance Dataset* measures the latent scores of organizational performance based on the 64 MOPAN evaluations of 32 IOs assessed between 2009 and 2021. Established as a non-partisan multilateral network of aid donors, MOPAN evaluates performance using a pre-defined assessment grid informed by documentary evidence and expert assessments (MOPAN, 2019). Under its current methodology, the MOPAN evaluation grid includes 12 key performance indicators (KPIs) in five areas, each underpinned by several micro-indicators (MIs). Using latent factor analysis, recent research has confirmed the presence of two dimensions of performance in the MOPAN assessment indicators, namely outcome performance and process performance (Reinsberg & Siauwijaya, 2023).

Our ultimate research data includes 48 IOs from 2000 to 2020 (Table A1 in the appendix). By including IOs that were not evaluated but which are similar in terms of their significance as development providers, we can construct a control group that we can observe to estimate counterfactual resourcing trends in the absence of performance evaluations. IOs evaluated by MOPAN include diverse organizations from ADB, AfDB, and CGIAR, to the World Bank, WFP, and WHO. Non-evaluated IOs include all remaining UN specialized agencies, development banks, and pass-through IOs for which core funding data were available, such as EBRD, IAEA, IMO,

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

and WIPO. We run balance tests to ensure that these are not different in fundamental ways compared to treated IOs (Table A2 in the appendix). By adding pre-treatment dummies, we can estimate trends in resourcing before MOPAN began to evaluate IO performance. Our dataset remains unbalanced due to missing data on core funding and when organizations were only created during the sample period.

3.2.2 Outcome variables

We construct two outcome measures that capture the quantity of funding received. We separately measure the amounts of *core funding* and *earmarked funding* that an organization receives from OECD/DAC donors in a given year. Both funding aggregates are included in the *Earmarked Funding Dataset*. We obtained the annual amount of earmarked funding by aggregating the activity-level contribution amounts from reporting donors to that organization in that year. We did not include activities that—according to the coding of the data—are not earmarked in any dimension but add those to the core contributions instead. As the official reporting quality has constantly improved, the figures are very similar to the official OECD/DAC reports.

For further tests, we compute the *stringency of earmarking* by taking the average of the activity-level stringency over all earmarked activities in an organization in a given year. Arguably, stringency is not directly observable, which suggests the use of latent factor analysis. We input the seven key indicators from the Earmarked Funding Data that capture earmarking across three dimensions. In each dimension, the data allows us to distinguish soft earmarking from hard earmarking. First, from the perspective of a global agency like the UN, activities can be earmarked by geography, either to specific regions (soft earmark), or specific countries (hard earmark). Second, activities can be earmarked by sector, either to specific themes (soft earmark) or specific sub-themes or circumscribed interventions (hard earmark). Third, earmarking may target specific branches, offices, or institutional sub-accounts (soft earmark), or require secondment from donor bureaucracies into the organization (hard earmark). While these earmarking dimensions and earmarking levels are theoretically derived, they may not correspond to earmarking stringency as perceived by agencies, which is why latent factor analysis is useful. Factor analysis with these seven indicators indicates that there is only one latent factor of earmarking stringency, which loads positively onto country-specific earmarking and project-type earmarking, but (weakly) negatively onto the remaining indicators (Table A3). We therefore retain one latent factor that corresponds to instances of stringent earmarking (Table A4).

3.2.3 Key predictors

Our main predictors capture two dimensions of performance. *Outcome performance*—our focal dimension of analysis—captures the extent to which organizations are perceived as delivering results on the ground. The variance contribution to this factor comes from four KPIs that capture results. *Process performance* captures the extent to which organizations are perceived as having the rules, systems, and procedures in place in order to manage operations effectively and efficiently, including strategic management, operational management, knowledge management, and rela-

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

tionship management. Factor analysis confirms the existence of these theoretically derived dimensions by showing that the MOPAN performance indicators load onto two latent dimensions (Reinsberg & Siauwijaya, 2023). The appendix provides further information about the MOPAN performance dataset (Box A1).

If we take donor rhetoric about results-based aid at face value, we would expect that donors primarily respond to outcome performance. Therefore, we use outcome performance as our main dimension and relegate tests using process performance to the appendix. Given our interest in establishing whether donors reward any good performance and punish any bad performance, we dichotomize the continuous performance measure by creating dummies indicating above-average performance ratings. In other words, our main dummy of interest is 'one' if an organization received a positive performance evaluation.

3.2.4 Research design

We perform two-way fixed effects analysis to examine whether performance information affects donor contributions. This analysis requires the parallel-trends assumption, which postulates that financial outcomes would have evolved similarly across treated IOs and untreated IOs. We believe this is a plausible assumption given our choice of control IOs which are similar to the evaluated IOs in important ways (Table A2). Our results would be biased under different pre-treatment trends, for example when donors were already determined to increase their financial commitments to evaluated IOs but decided to conduct an evaluation to justify their policy choices. We probe the possibility of diverging trends by including dummies for each year in the five-year period up to an evaluation.

Confounding bias is also unlikely because the MOPAN board—composed of member state representatives—decides whether an organization will be evaluated but the timing of MOPAN evaluations appears to follow a relatively rigid schedule. IOs get regularly evaluated, about every four years. Sometimes, it may take up to five years until an organization is evaluated again, depending on available capacities determined by the number of other IOs evaluated in the same assessment cycle. As confirmed by interviews with former MOPAN officials, it is unlikely that member states exercise control over this process to a point where the decision to evaluate an organization would be endogenous to their financial contributions. Therefore, we believe that threats to internal validity are minimal. That said, however, the set of evaluated IOs is not representative of all IOs. The MOPAN methodology states that evaluations cover the most important IOs that are also likely to receive contributions from DAC members. In choosing our set of control IOs, we sought to mirror this criterion.

Two-way fixed-effects estimators make other assumptions that may be problematic for our inferences (de Chaisemartin & D'Haultfoeuille, 2023). One is the absence of carry-over effects, which states that current outcomes do not depend on past treatments. This assumption cannot be violated where we study the long-term impact of the first performance assessment of an organization, ignoring all future assessments (Borusyak et al., 2024). Alternatively, to allow for treatment reversal, we may formulate an additional assumption that responses to new performance information will

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

fade off after a few years, typically until the next assessment. This implies a different estimator, which we also present below (Imai et al., 2023).

A final assumption is that treatment effects are homogeneous, as reflected in our choice of a binary treatment indicator. However, treatment may be continuous. Different performance scores reflect different treatment doses. Identification of such designs is challenging because lower-dose cases may not be valid counterfactuals for higher-dose cases (Callaway et al., 2021). One might think such concerns are less relevant in our case because IOs have incentives to obtain high assessment scores, thereby mitigating the problem of self-selection into higher-dose treatments. However, we follow a more conservative approach by grouping IOs into a discrete number of bins in which performance scores can be regarded as reflecting similar actual performance.

In summary, we perform variants of two-way fixed-effects estimations controlling for unobserved time-invariant IO variables like headquarter location and common shocks to all IOs like geopolitical shifts. The appendix shows variable definitions and descriptive statistics for all variables (Table A5).

3.3 Findings

Figure 6 presents our key results on the relationship between outcome performance and donor contributions differentiated by funding modality. We find a consistently positive relationship between a positive evaluation of performance and the receipt of earmarked funding in most post-treatment years in the five-year horizon after the evaluation. Substantively, a positive evaluation promises additional contributions of about 49.2% (95%-CI: 2.6-116.8%) after three years. The cumulative amount of earmarked contributions after five years is higher. In contrast, we do not find any

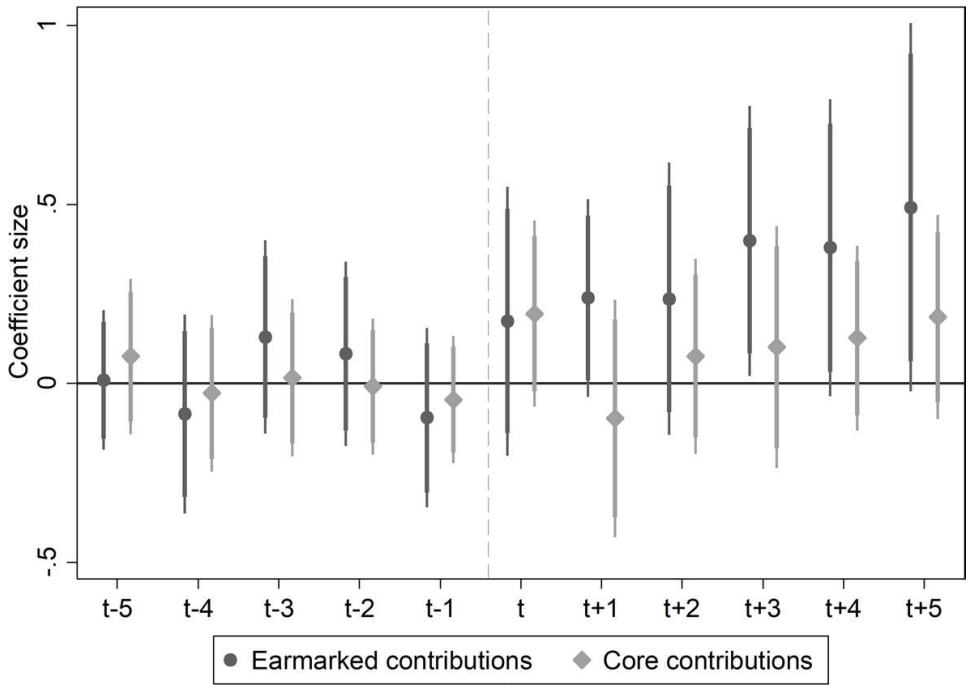


Fig. 6 Positive outcome performance evaluation and donor contributions to IOs. *Notes:* Thin whiskers show the 95%-CI, bold whiskers show the 90%-CI

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

relationship between performance and core contributions. This underscores the importance of disaggregating funding streams. The appendix presents full coefficient tables for donor contributions following both positive outcome performance (Table A6) and process performance (Table A7).

In the appendix, we probe the robustness of these findings further. Specifically, we explicitly control for negative evaluations, which changes the interpretation of the main coefficients. We find that a positive evaluation has a significantly positive impact on earmarked contributions relative to the funding evolution of non-assessed organizations. We do not find evidence of similar effects for core contributions (Figure A1).

In addition, we seek to capture heterogeneity within positive performance assessments. Given the strong identification assumptions of continuous-treatment models, we deploy a binning estimator whereby we group positive performance scores in three bins. We cut the sample after each standard deviation of the continuous score. The results show that the positive relationship between outcome performance and earmarked funding is driven by positive assessments up to one standard deviation (SD) above the mean and particularly above 2 SDs (Table A8). The middle bin—between 1 SD and 2 SDs—shows negatively significant coefficients for both funding types, which is puzzling and might be explained by composition effects.[7]

We are confident that our results are not driven by different pre-treatment trends in the outcome variable because the pre-treatment trends are statistically insignificant. To further mitigate pre-treatment trends as potential confounder, we employ alternative estimators in robustness tests. Specifically, the PanelMatch estimator only keeps control units with the same treatment history as the treated units for a consecutive number of pre-treatment years specified by the researcher. Furthermore, the estimator uses propensity-score weighting to ensure that treated units have similar pre-treatment outcomes as the control units in the pre-treatment years (Imai et al., 2023). Identification assumes unconfoundedness of potential outcomes up to the chosen number of years in which units must have identical treatment histories. The choice of this lag number is arbitrary but should be made based on knowledge of context. Given that the modal waiting time for consecutive MOPAN evaluations is about four years, we choose a pre-treatment lag of six years. Furthermore, we compute observation weights that ensure similarity of pre-treatment outcomes up to three years prior to the treatment. We find a consistently positive effect of a positive evaluation of outcome performance with respect to earmarked contributions after three years (Figure A2). The point estimate translates into a funding increase by about 55.2% (95%-CI: 4.6-129.3%).

In addition, we use a different estimator that allows us to also probe the long-term benefits of a positive evaluation. For this estimator, treatment can be staggered but is assumed to be an absorbing state (Borusyak et al., 2024). In our data, this implies that the treatment indicator is one from the year of a positive outcome performance

[7] In fact, all four IOs in this bin are MDBs. This robustness tests thus illustrates the problems with multi-valued treatments in two-way fixed-effects regression where negative weights may bias the estimates of interest. Using a dichotomized treatment (as we did in our main analysis)—the advice of methodologists—avoids negative weights (Callaway et al., 2021, p. 26).

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

evaluation until the final sample year. This disregards the fact that organizations can receive negative evaluations in subsequent evaluations. To mitigate spill-over effects from subsequent evaluations, we limit the plotting horizon to six years after the first evaluation. We find significantly positive effects of a positive evaluation with regard to earmarked contributions (Figure A3).

Demonstrating the versatility of the Earmarked Funding Dataset, we now explore donor heterogeneity in resourcing responses to performance evaluations. To that end, we use a more disaggregated component of our dataset that records the earmarked contributions from individual donors to individual organizations. Using this fine-grained dependent variable, we can estimate how a given donor responds to a positive evaluation of outcome performance. We would expect stronger responses from MOPAN members, given their commitment to effective multilateralism and performance-based aid allocation. Table A9 in the appendix shows descriptive statistics for the disaggregated dataset.

Table 1 presents the results. Using the previous set of fixed effects, we find that MOPAN members significantly increase their earmarked contributions after a positive evaluation of outcome performance. Non-MOPAN donors appear to reduce their contributions, but this association vanishes once we only consider within-donor variation in the second model. In the third model, we control for arbitrary donor-year effects, which eliminates the confounding effect of changing aid priorities, donor budgets, and other donor-specific shocks. The differences across donor types in their responses to performance shocks decrease, although effect magnitudes continue to be significantly larger for MOPAN members. For example, five years after the evaluation, non-MOPAN donors would be predicted to increase their contributions from $1.52 million to $1.67 million if the evaluation was positive. For MOPAN members, the average increase would be from $1.81 million to $2.94 million.

Finally, we illustrate the usefulness of our dataset by probing earmarking stringency. We test whether positive outcome performance evaluations are associated with less stringency of earmarking. We apply a weighted average to the activity-specific latent factor of earmarking stringency where weights are proportional to the project amount. We run the analysis pooling all donors together and then allow for effect heterogeneity according to whether a donor is a MOPAN member. Table 2 shows the results. We do not find evidence that donors reduce the stringency of earmarking following a positive evaluation of outcome performance. To the contrary, earmarking appears to slightly increase in the third year after the evaluation, indicating that additional fundraising following such performance information may be decentralized. In the disaggregated analysis, we do not find evidence for any response by any donor. However, it appears that MOPAN members are unconditionally less stringent donors with respect to earmarking. In sum, these results imply that donors give more funding following positive performance evaluations and that these increased funds do not come at a cost of decreased discretion for IOs.

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

Table 1 Individual donor responses to positive evaluation of outcome performance

	Total contributions		Earmarked contributions		Core contributions	
	(1)		(2)		(3)	
Positive outcome performance (non-members)						
t	-0.061	(0.051)	0.018	(0.035)	0.076**	(0.033)
t+1	-0.090**	(0.042)	-0.016	(0.029)	0.052**	(0.025)
t+2	-0.163***	(0.061)	-0.028	(0.046)	0.056	(0.040)
t+3	-0.135	(0.088)	0.011	(0.070)	0.086	(0.062)
t+4	-0.143	(0.097)	0.002	(0.077)	0.097	(0.069)
t+5	-0.171	(0.124)	0.092	(0.060)	0.165***	(0.051)
Additional effects (MOPAN members)						
MOPAN member	0.495***	(0.089)	0.156***	(0.031)	---	---
t × MOPAN	0.325**	(0.146)	0.187**	(0.079)	0.085	(0.074)
t+1 × MOPAN	0.344**	(0.144)	0.216***	(0.070)	0.100	(0.069)
t+2 × MOPAN	0.609***	(0.153)	0.375***	(0.070)	0.230***	(0.071)
t+3 × MOPAN	0.550***	(0.177)	0.301***	(0.093)	0.171*	(0.095)
t+4 × MOPAN	0.566**	(0.221)	0.315**	(0.151)	0.151	(0.144)
t+5 × MOPAN	0.833***	(0.283)	0.388**	(0.181)	0.263	(0.171)
Pre-treatment trends						
t-1	0.144***	(0.034)	0.144***	(0.035)	0.144***	(0.035)
t-2	0.147***	(0.042)	0.146***	(0.042)	0.146***	(0.043)
t-3	0.115**	(0.046)	0.116**	(0.046)	0.116**	(0.046)
t-4	0.040	(0.036)	0.040	(0.036)	0.041	(0.036)
t-5	0.004	(0.038)	0.004	(0.038)	0.004	(0.038)
IO-fixed effects	X					
Dyad-fixed effects			X		X	
Year-fixed effects	X		X			
Donor-year fixed effects					X	
Diagnostics						
Observations	30,240		30,240		30,240	
Clusters	48		48		48	
R-squared	0.351		0.662		0.700	

Notes OLS regression with fixed effects as indicated and robust standard errors clustered on IOs in parenthesis. Significance levels: * $p<.1$ ** $p<.05$ *** $p<.01$

4 Summary and outlook

We introduced the Earmarked Funding Dataverse—a collection of datasets that allow researchers to comprehensively track the finances of 340 IOs in development from 1990 to 2020 based on a systematic coding of 342,812 earmarked activities from 50 official donors. Compared to existing related datasets, the Earmarked Funding Dataset provides longer temporal coverage and assesses the stringency of earmarking at the activity level. These features facilitate empirical tests of theoretical propositions in the related literature concerning when and how donors use money to wield influence in contemporary IOs. They also promise to promote research on the effectiveness of IOs.

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

Table 2 Positive outcome performance evaluation and earmarking stringency

	Latent earmarking index			
	(1)		(2)	
Positive outcome performance				
t	-0.014	(0.114)	0.006	(0.015)
t+1	0.120	(0.103)	-0.002	(0.009)
t+2	0.064	(0.127)	-0.008	(0.008)
t+3	0.254*	(0.139)	0.013	(0.015)
t+4	0.097	(0.109)	-0.001	(0.011)
t+5	0.297	(0.182)	-0.006	(0.011)
Pre-treatment trends				
t-1	-0.093	(0.069)	-0.011**	(0.005)
t-2	0.045	(0.073)	-0.001	(0.006)
t-3	-0.028	(0.073)	-0.006	(0.007)
t-4	0.032	(0.076)	-0.003	(0.007)
t-5	0.049	(0.073)	0.005	(0.006)
Additional variables				
MOPAN member			-0.039***	(0.006)
t × MOPAN			-0.005	(0.016)
t+1 × MOPAN			0.006	(0.010)
t+2 × MOPAN			0.015	(0.009)
t+3 × MOPAN			-0.004	(0.014)
t+4 × MOPAN			0.006	(0.017)
t+5 × MOPAN			0.022	(0.015)
Diagnostics				
Observations	873		8666	
Clusters	48		48	
R-squared	0.195		0.322	

Notes OLS regression with two-way fixed effects. In the donor-specific regressions, fixed effects refer to dyads. Robust standard errors clustered on IOs in parenthesis. Significance levels: * $p<.1$ ** $p<.05$ *** $p<.01$

We first examined the data descriptively to establish the key trends in the evolution of the financial architecture of development cooperation through IOs. We found that toward the turn of the Millennium, states have created new development IOs as pass-through IOs, like the Global Environment Facility, the Global Fund, and the Global Partnership on Education. Concurrently, IOs have increasingly established institutional sub-accounts, specifically special-purpose trust funds, to receive earmarked contributions from their donors. Earmarked funding has become a significant source of funding. An increasing portion of such earmarked aid for operational IOs has been coming from pass-through IOs. Finally, within earmarked resources, the stringency of earmarking has remained relatively stable, despite an increase in the amount earmarked funding in the least-earmarked category from a low base in recent years.

As a first empirical application using the data, we examined whether donors reward organizations that demonstrate positive performance in MOPAN assessments. In doing so, we revisited and extended earlier research showing that donors do not respond to performance consistently but only do so if agencies provide services that make them relatively exchangeable and if agencies do not have strong alliances with non-state actors (Lall, 2021). Our contribution was to unpack different financial flows, distinguishing between performance rewards through core funding and earmarked funding, as well as earmarking stringency. Using two-way-fixed-effects

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

estimations, we found that donors respond to performance evaluations by adjusting earmarked funding but not core funding. This finding is in line with our theoretical expectations as earmarked funding is much more flexible from a donor perspective than core contributions which are subject to multi-year plans and (implicit) burden-sharing arrangements with other donors. We further demonstrated that funding is most sensitive to performance among MOPAN member states. Finally, we derived a latent index of earmarking stringency and found that donors do not generally adjust their earmarking stringency in response to performance information. Together, these findings imply that donors reward well-performing IOs by increasing their funds and that these increased funds do not come at a cost of lower discretion for IOs.

We note the limitations of our data collection and empirical analysis while pointing out avenues for future research. Our dataset includes only IOs that are eligible to receive ODA resources to support activities in development. For these types of IOs, financial resources are important to carry out their operations. Moreover, our dataset only records contributions from donors that report aid flows to the DAC, which excludes non-DAC donors like China. While researchers have collected data on Chinese development finance (Custer et al., 2021), such data do not include Chinese contributions to development IOs. Future research should therefore collect this data. Due to the omission of new donors as well as under-reporting in DAC statistics in earlier years, we are unlikely to capture all earmarked funding. The figures presented in this article should therefore be understood as a lower bound of earmarked funding in the multilateral system. Future research could improve the coverage for earlier years by triangulating the data with alternative sources, such as the annual budget reports of IOs.

Data coverage issues also beclouded the inferences in our analysis of performance assessments and funding choices. Despite the expanded sample afforded by the earmarked funding data, our results are based on short time series, which complicates robust inferences due to potential problems with statistical power and asymptotic assumptions in the estimators. A particular challenge is the short time series for core funding. Future research should try to collect prolonged funding data to facilitate the use of matching techniques that would bolster our confidence in the results. Causal identification rests on ultimately untestable assumptions, notably parallel trends between treatment group and control group. Theory thus has a greater role of play to develop observable implications that could help untangle the mechanisms underlying the relationship between performance assessments and funding choices. There is also scope to further explore effect heterogeneity, including through qualitative case studies, to inform theory development.

We hope the *Earmarked Funding Dataset* will help scholars extend the frontier of comparative-quantitative research on international organizations and development policy. Descriptively, the data will promote a better understanding of how contemporary development IOs are financed. In fact, official statistics often underestimate the true amount of development finance that is earmarked for specific purposes, by counting contributions to pass-through IOs at their source rather than as earmarked outflows to operational IOs. Beyond descriptive purposes, the data can be used to facilitate research on the effectiveness of development IOs at different levels of analysis. What makes aid projects successful? Why are some IOs more effective than

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

others in contributing to the Sustainable Development Goals? Why do some organizational units in the same IO perform better than others? The combined insights from these levels of analysis may help scholars to gain a more complete picture of when and why aid is effective. Finally, the data can be used to examine the drivers of earmarked funding. Why do some donors earmark more than others? Why are some IOs more susceptible to earmarking than others? Why do some donors earmark some IOs more than they earmark others? And how do IOs respond to earmarking pressures? With the *Earmarked Funding Dataset*, debates about these questions can become even more insightful as researchers now have better data at hand to test their predictions empirically.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11558-024-09548-1.

Acknowledgements We thank participants of the Global Policy North workshop in Glasgow (24 February 2023) for valuable comments. We are grateful to Emily Bosch, Abdoulaye Fabregas, and Jieun Kim from the OECD Development Cooperation Directorate for valuable feedback and constructive exchanges. Funding from a UK Research and Innovation Future Leaders Fellowship grant (MR/V022148/1) is gratefully acknowledged.

Data availability The Earmarked Funding Data are freely available on Harvard Dataverse (https://doi.org/10.7910/DVN/WCS99H) and the *Review of International Organizations*' webpage.

Declarations

Conflict of interest The authors declare no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Abbott, K. W., & Snidal, D. (1998). Why states act through formal international organizations. *Journal of Conflict Resolution*, *42*(1), 3–32. https://doi.org/10.1177/0022002798042001001.

Barder, O., Ritchie, E., & Rogerson, A. (2019). Contractors or Collectives? Earmarked Funding of Multilaterals, Donor Needs, and Institutional Integrity: The World Bank as a Case Study. *CGD Policy Paper*, *153*(July). Washington D.C.: Center for Global Development.

Baumann, M. O., & Weinlich, S. (2020). Funding the UN: Support or constraint? In S. Browne, & T. G. Weiss (Eds.), *Routledge Handbook on the UN and Development funding the UN support or constraint?* (pp. 151–164). Taylor & Francis Group.

Baumann, M. O., Lundsgaarde, E., & Weinlich, S. (2019). Shades of grey: Earmarking in the UN development system. *Financing the UN Development System: Time for hard choices* (pp. 106–109). UNDP.

Bayram, A. B., & Graham, E. R. (2022). Knowing how to give: International Organization Funding Knowledge and Public Support for Aid Delivery channels. *Journal of Politics*, *84*(4), 1885–1898.

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

Bermeo, S. B. (2017). Aid allocation and targeted development in an increasingly connected World. *International Organization*, *71*(4), 735–766. https://doi.org/10.1017/S0020818317000315.

Borusyak, K., Jaravel, X., & Spiess, J. (2024). Revisiting event study designs: Robust and efficient estimation. *Review of Economic Studies*. https://doi.org/10.1093/restud/rdae007.

Broome, A., Homolar, A., & Kranke, M. (2018). Bad science: International organizations and the indirect power of global benchmarking. *European Journal of International Relations*, *24*(3), 514–539. https://doi.org/10.1177/1354066117719320.

Brutger, R., & Clark, R. (2022). At what cost? Power, payments, and Public Support of International Organizations. *Review of International Organizations*, *18*, 431–465.

Callaway, B., Goodman-Bacon, A., & Sant'Anna, P. H. C. (2021). Difference-in-differences with a continuous treatment. Arxiv:2107.02637v4. Accessed 1 February 2024.

Conceição-Heldt, E. (2017). Regaining control of errant agents? Agency slack at the European Commission and the World Health Organization. *Cooperation and Conflict*, *52*(4), 469–484.

Custer, S., Dreher, A., Elston, T., Fuchs, A., Ghose, S., Lin, J. J., Malik, A., Parks, B. C., Russell, B., Solomon, K., Strange, A., Tierney, M. J., Walsh, K., Zaleski, L., & Zhang, S. (2021). *Tracking Chinese Development Finance: An Application of AidData's TUFF 2.0 Methodology*. *AidData Working Paper*. Williamsburg: AidData at William & Mary.

de Chaisemartin, C., & D'Haultfoeuille, X. (2023). *Difference-in-differences for simple and complex natural experiments*. Princeton University Press.

De Vries, C. E., Hobolt, S. B., & Walter, S. (2021). Politicizing international cooperation: The mass public, political entrepreneurs, and political opportunity structures. *International Organization*, *75*(2), 306–332.

Dellmuth, L. M., Scholte, J. A., Tallberg, J., & Verhaegen, S. (2022). *Citizens, elites, and the legitimacy of global governance*. Oxford University Press.

Dietrich, S. (2021). *States, Markets, and Foreign Aid*. Cambridge University Press.

Dijkstra, H. (2014). Shadow bureaucracies and the unilateral control of international secretariats: Insights from UN peacekeeping. *Review of International Organizations*, *10*(1), 23–41. https://doi.org/10.1007/s11558-014-9203-7.

Eckhard, S., Jankauskas, V., Leuschner, E., Burton, I., Kerl, T., & Sevastjanova, R. (2023). Natural language processing of evaluation reports: Introducting a novel performance metric for international organizations. *Review of International Organizations*. https://doi.org/10.1007/s11558-023-09489-1.

Eichenauer, V. Z., & Reinsberg, B. (2017). What determines earmarked funding to international development organizations? Evidence from the new multi-bi aid data. *Review of International Organizations*, *12*(2), 171–197. https://doi.org/10.1007/s11558-017-9267-2.

Fleck, R. K., & Kilby, C. (2006). How do political changes influence US bilateral aid allocations? Evidence from panel data. *Review of Development Economics*, *10*(2), 210–226.

Graham, E. R. (2017). Follow the money: How trends in Financing are changing governance at International Organizations. *Global Policy*, *8*(August), 15–25. https://doi.org/10.1111/1758-5899.12450.

Graham, E. R. (2023). *Transforming International Institutions: How money quietly sidelined multilateralism at the United Nations*. Oxford University Press.

Hawkins, D. G., Lake, D. A., Nielson, D. L., & Tierney, M. J. (2006). *Delegation under anarchy: States, international organizations, and principal-agent theory*. Cambridge University Press.

Heinzel, M., Cormier, B., & Reinsberg, B. (2023). Earmarked funding and the control–performance tradeoff in international development organizations. *International Organization*, *77*(2), 475–495.

Heinzel, M., & Reinsberg, B. (2024). Trust funds and the sub-national effectiveness of development aid: Evidence from the World Bank. *World Development,* 179, 106609. https://doi.org/10.1016/j.worlddev.2024.106609

Hirschmann, G. (2021). International organizations' responses to member state contestation: From inertia to resilience. *International Affairs*, *97*(6), 1963–1981.

Honig, D., & Weaver, C. (2019). A race to the top? The aid transparency index and the Social Power of Global Performance indicators. *International Organization*, *73*(3), 579–610. https://doi.org/10.1017/S0020818319000122.

Hooghe, L., & Marks, G. (2015). Delegation and pooling in international organizations. *Review of International Organizations*, *10*(3), 305–328. https://doi.org/10.1007/s11558-014-9194-4.

Hüfner, K. (2017). The financial crisis of UNESCO after 2011: Political reactions and organizational consequences. *Global Policy*, *8*(5), 96–101.

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

Imai, K., Kim, I. S., & Wang, E. H. (2023). Matching methods for causal inference with Time-Series Cross-sectional Data. *American Journal of Political Science*, *67*(3), 587–605. https://doi.org/10.1111/ajps.12685.

Jones, S., & Tarp, F. (2016). Does foreign aid harm political institutions? *Journal of Development Economics*, *118*, 266–281.

Kelley, J. G., & Simmons, B. A. (2015). Politics by number: Indicators as social pressure in international relations. *American Journal of Political Science*, *59*(1), 55–70. https://doi.org/10.1111/ajps.12119.

Lall, R. (2021). The Financial consequences of Rating International Institutions: Competition, collaboration, and the politics of Assessment. *International Studies Quarterly*, *65*(2), 343–359. https://doi.org/10.1093/isq/sqaa082.

Lall, R. (2023). Making global governance Accountable: Civil Society, States, and the politics of reform. *American Journal of Political Science*. https://doi.org/10.1111/ajps.12824.

Lundgren, M., Tallberg, J., Sommerer, T., & Squatrito, T. (2023). When are international organizations responsive to policy problems? *International Studies Quarterly*, *67*(3), sqad045.

Michaelowa, K., Reinsberg, B., & Schneider, C. (2018). The politics of delegation in the European Union. *International Studies Quarterly*. https://doi.org/10.1093/isq/sqy034.

Milner, H. V., & Tingley, D. (2013). The choice for multilateralism: Foreign aid and American foreign policy. *Review of International Organizations*, *8*, 313–341.

MOPAN (2019). *MOPAN 3.0 Methodology Manual*. Paris. http://www.mopanonline.org/ourwork/ourapproachmopan30/Methodology Manual- FINAL.pdf. Paris: Multilateral Organisation Performance Assessment Network.

Nielson, D. L., & Tierney, M. J. (2003). Delegation to International organizations: Agency Theory and World Bank Environmental Reform. *International Organization*, *57*(02), 241–276.

OECD. (2012). *2012 Multilateral Aid Report. Paris: Organization for Economic Cooperation and Development*. Organization for Economic Cooperation and Development.

OECD. (2020). *Earmarked funding to multilateral organisations: How is it used and what constitutes good practice?* Organization for Economic Cooperation and Development.

OECD (2021). Creditor Reporting System (CRS). stats.oecd.org. Accessed 20 October 2021.

Palagashvili, L., & Williamson, C. R. (2021). Grading foreign aid agencies: Best practices across traditional and emerging donors. *Review of Development Economics*, *25*(2), 654–676. https://doi.org/10.1111/rode.12747.

Passmore, T., Shannon, M., & Nadeau, M. (2023). Financial contributions to United Nations Peacekeeping, 1990–2010: A new dataset. *Conflict Management and Peace Science*, *40*(1), 88–107.

Patz, R., & Goetz, K. H. (2019). *Managing money and discord in the UN: Budgeting and bureaucracy*. Oxford University Press.

Pevehouse, J. C., Nordstrom, T., McManus, R., & Jamison, A. S. (2021). Tracking organizations in the world: The correlates of War IGO Version 3.0 datasets. *Journal of Peace Research*, *57*(3), 492–503.

Reinsberg, B. (2016). The implications of multi-bi financing on multilateral agencies: The example of the World Bank. In T. Mahn, M. Negre, & S. Klingebiel (Eds.), *The fragmentation of aid: Concepts, measurements and implications for development cooperation* (pp. 185–198). Palgrave Macmillan.

Reinsberg, B. (2023). Earmarked funding and the performance of international organizations: Evidence from food and agricultural development agencies. *Global Studies Quarterly*, *3*(4), ksad056. https://doi.org/10.1093/isagsq/ksad056

Reinsberg, B. (2017). Organizational reform and the rise of trust funds: Lessons from the World Bank. *Review of International Organizations*, *12*(2), 199–226. https://doi.org/10.1007/s11558-017-9268-1.

Reinsberg, B., & Siauwijaya, C. (2024). Does earmarked funding affect the performance of international organizations? *Global Policy*, 15(1), 23–39. https://doi.org/10.1111/1758-5899.13270.

Reinsberg, B., Michaelowa, K., & Eichenauer, V. Z. (2015). The rise of multi-bi aid and the proliferation of trust funds. In M. Arvin, & B. L. Lew (Eds.), *Handbook on the Economics of Foreign Aid* (pp. 527–554). Edward Elgar.

Reinsberg, B., Michaelowa, K., & Knack, S. (2017). Which Donors, which funds? Bilateral donors' choice of Multilateral funds at the World Bank. *International Organization*, *71*(4), 767–802. https://doi.org/10.1017/S0020818317000340.

Reykers, Y., & Adriaensen, J. (2022). The politics of understaffing international organisations: The EU Military Planning and Conduct Capability (MPCC). *European Security*, *0*(0), 1–20. https://doi.org/10.1080/09662839.2022.2142040.

Rodrik, D. (1995). *Why is there multilateral lending?* NBER Working Paper 5160.

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

Schmid, L. K., Reitzenstein, A., & Hall, N. (2021). Blessing or curse? *Global Governance: A Review of Multilateralism and International Organizations*, *27*(3), 433–459. https://doi.org/10.1163/19426720-02703002

Schneider, C. J., & Tobin, J. L. (2016). Portfolio similarity and international development aid. *International Studies Quarterly*, *60*(4), 647–664. https://doi.org/10.1093/isq/sqw037.

Sridhar, D., & Woods, N. (2013). Trojan Multilateralism: Global Cooperation in Health. *Global Policy*, *4*(4), 325–335.

Thorvaldsdottir, S., & Patz, R. (2019). IO bureaucracies between Resource mobilization and Problem-Solving: The case of the UNHCR. *International Review of Administrative Sciences*, *87*(4), 794–812.

UN-MPTF. (2019). *Financing the UN Development System: Time for hard choices*. Dag Hammerskjold Foundation.

United Nations (2022). UN Data Cube. https://unsceb.org/data-standards-united-nations-system-wide-reporting-financial-data. Accessed 1 May 2022.

Von Borzyskowski, I., & Vabulas, F. (2019). Hello, goodbye: When do states withdraw from international organizations? *Review of International Organizations*, *14*(2), 335–366.

Weaver, C. (2007). The World's Bank and the Bank's World. *Global Governance*, *13*(4), 493–512. http://www.jstor.org/stable/27800679.

Zürn, M., Tokhi, A., & Binder, M. (2021). The international authority database. *Global Policy*, *12*(4), 430–442.

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

Authors and Affiliations

Bernhard Reinsberg[1,2] · Mirko Heinzel[1,3] · Christian Siauwijaya[1]

✉ Bernhard Reinsberg
bernhard.reinsberg@web.de; Bernhard.Reinsberg@glasgow.ac.uk

[1] School of Social and Political Sciences, University of Glasgow, Adam Smith Building, Glasgow G12 8RT, United Kingdom

[2] Centre for Business Research, University of Cambridge, Cambridge, United Kingdom

[3] Faculty of Arts and Social Sciences, Maastricht University, Maastricht, Netherlands

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

Terms and Conditions

Springer Nature journal content, brought to you courtesy of Springer Nature Customer Service Center GmbH ("Springer Nature").

Springer Nature supports a reasonable amount of sharing of research papers by authors, subscribers and authorised users ("Users"), for small-scale personal, non-commercial use provided that all copyright, trade and service marks and other proprietary notices are maintained. By accessing, sharing, receiving or otherwise using the Springer Nature journal content you agree to these terms of use ("Terms"). For these purposes, Springer Nature considers academic use (by researchers and students) to be non-commercial.

These Terms are supplementary and will apply in addition to any applicable website terms and conditions, a relevant site licence or a personal subscription. These Terms will prevail over any conflict or ambiguity with regards to the relevant terms, a site licence or a personal subscription (to the extent of the conflict or ambiguity only). For Creative Commons-licensed articles, the terms of the Creative Commons license used will apply.

We collect and use personal data to provide access to the Springer Nature journal content. We may also use these personal data internally within ResearchGate and Springer Nature and as agreed share it, in an anonymised way, for purposes of tracking, analysis and reporting. We will not otherwise disclose your personal data outside the ResearchGate or the Springer Nature group of companies unless we have your permission as detailed in the Privacy Policy.

While Users may use the Springer Nature journal content for small scale, personal non-commercial use, it is important to note that Users may not:

1. use such content for the purpose of providing other users with access on a regular or large scale basis or as a means to circumvent access control;
2. use such content where to do so would be considered a criminal or statutory offence in any jurisdiction, or gives rise to civil liability, or is otherwise unlawful;
3. falsely or misleadingly imply or suggest endorsement, approval , sponsorship, or association unless explicitly agreed to by Springer Nature in writing;
4. use bots or other automated methods to access the content or redirect messages
5. override any security feature or exclusionary protocol; or
6. share the content in order to create substitute for Springer Nature products or services or a systematic database of Springer Nature journal content.

In line with the restriction against commercial use, Springer Nature does not permit the creation of a product or service that creates revenue, royalties, rent or income from our content or its inclusion as part of a paid for service or for other commercial gain. Springer Nature journal content cannot be used for inter-library loans and librarians may not upload Springer Nature journal content on a large scale into their, or any other, institutional repository.

These terms of use are reviewed regularly and may be amended at any time. Springer Nature is not obligated to publish any information or content on this website and may remove it or features or functionality at our sole discretion, at any time with or without notice. Springer Nature may revoke this licence to you at any time and remove access to any copies of the Springer Nature journal content which have been saved.

To the fullest extent permitted by law, Springer Nature makes no warranties, representations or guarantees to Users, either express or implied with respect to the Springer nature journal content and all parties disclaim and waive any implied warranties or warranties imposed by law, including merchantability or fitness for any particular purpose.

Please note that these rights do not automatically extend to content, data or other material published by Springer Nature that may be licensed from third parties.

If you would like to use or distribute our Springer Nature journal content to a wider audience or on a regular basis or in any other manner not expressly permitted by these Terms, please contact Springer Nature at

onlineservice@springernature.com