SUERF
The European Money and Finance Forum

SUERF Policy Note
Issue No 284, August 2022

The design of a data governance system[1]

By Siddharth Tiwari*, Sharad Sharma**, Siddharth Shetty**, and Frank Packer*

JEL codes: G28, H41, K00, L17, L51, O33

Keywords: Data governance, big techs, data-sharing, data taxonomy, consent framework, account aggregators, General Data Protection Regulation (GDPR), Data Empowerment Protection Architecture (DEPA).

Consumers often do not know the benefits of the data they generate and find it difficult to assert their rights regarding the collection, processing and sharing of their data. We propose a data governance system that restores control to the parties generating the data, by requiring consent prior to their use by service providers. The system should be open, with consent that is revocable, granular, auditable, and with notice in a secure environment. Conditions also include purpose and use limitation, data minimisation, and retention restriction. Trust in the system and widespread adoption are enhanced by mandating specialised data fiduciaries. The experience with India's Data Empowerment Protection Architecture (DEPA) suggests that such a system can operate at scale with low transaction costs.

[1]Disclaimer: This is an abridged version of a much longer paper published as BIS Papers No. 124, May 2022. The views expressed in this paper are those of the authors and do not necessarily reflect the views of the BIS. We are indebted to Rahul Matthan for close collaboration on data governance issues that has imprinted this paper. We thank Agustín Carstens and Nandan Nilekani for discussions on these (and related) issues over the years. We are grateful to Sanjay Jain, Saurabh Panjwani, Vamsi Madhav, B G Mahesh, Pramod Varma, Raphael Auer, Stijn Claessens, Jon Frost and Hyun Song Shin as well as participants at a BIS seminar for their comments. Several conversations with Derryl D'Silva are also acknowledged. Shreya Ramann and Jenny Hung provided excellent support. All errors are ours.

*Bank for International Settlements

**iSPIRIT Foundation

1. Introduction

Throughout history, consumers and businesses have generated data through their everyday choices. These data could relate, inter alia, to doctor's visits, purchases and sales of goods, or financial transactions. Traditionally, this information was paper-based and resided with the entities that engaged in these transactions (doctors, merchants and financial service providers).

Technological developments over the last two decades have led to an explosion in the availability of data and their processing. The combination of the increased availability of data and inexpensive storage has provided the foundations for high-performance computation. It has also enabled the harnessing of very large amounts of consumer data – often referred to as "big data" – into a valuable commodity. In such a setting, the key questions are who has control over these data, where it is stored, with whom and under what conditions it is shared, and who operates the data governance system.

In most countries, privacy laws have enabled countries to create accompanying legislation that recognises the rights of individuals over their data. Central to these laws is a set of principles that define how personal data are collected, shared, and processed. However, in spite of these laws, generators of data such as consumers and small and medium-sized enterprises (SMEs) do not have control over the data they generate and are denied the opportunity to reap the full value from their use.

This is for a few reasons. First, a service provider usually seeks consent to use and transfer data at the time when a consumer agrees to participate in an activity with the service provider. Since this consent is sought ex ante and for a wide range of possibilities, it tends to be broad and sweeping in nature. Consumers – impatient and possibly ignorant of the value of the data they generate – quickly grant consent to gain access to the service. Second, newly created data are often gathered and retained in proprietary silos and stored in various institutions in incompatible formats. Even when aware of their value, consumers find it difficult to access their own data as they are in different formats and in different locations, and consumers have only limited options for combining data requests across institutions.

This paper proposes a data governance system that corrects for the above-mentioned market failures by restoring control of data to consumers and merchants generating the data – whom we refer to as data subjects. The system allows data subjects to effectively operationalise their rights with regard to the collection, processing and sharing of their data and requires service providers such as social media channels or lenders – whom we refer to as data users – to always provide notice and seek the consent of data subjects prior to sharing and processing their data. Such a consent system would replace "broad and sweeping" consent with "granular" consent. Such a consent-based system will empower data subjects to use their data for their own benefit.

The proposed data governance system is grounded in current national privacy laws in various jurisdictions that set out core data protection principles. The governance system describes how these privacy principles are operationalised. Given the granularity of data-sharing requests, the enormous amounts of data involved, the possibility of data being spread over several data users and providers, and the need to keep the data secure and cost-effective, consent-based systems with the above characteristics must be digital to meet these objectives. For a digital system to operate effectively, it should embody the protocols that translate the privacy framework to the digital space. This will include elements from principles of notice and consent, purpose limitation, data minimisation, retention restriction and use limitation. It will also need to be open, with consent that is revocable, granular, subject to audit, and with notice in a secure environment.

In the present situation, where data subjects are at a significant handicap, trust in the consent-based system as well as its widespread adoption has the potential to be significantly enhanced by mandating specialised data fiduciaries whose primary task – as advocates of data subjects – is to ensure that data are shared in a fashion that respects the above-mentioned principles of effective data governance. The experience with India's Data Empowerment Protection Architecture (DEPA) suggests that such a consent-based system can operate at scale with low transaction costs.

2. Data-sharing: participants and data taxonomy

There are four participants in the proposed data governance system:

Data subjects: individuals, consumers, and businesses whose activities (online or physical) generate data, and to whom the personal data pertain.

Data providers: entities where data are stored, often also referred to as data controllers. These entities are service providers – such as financial institutions, big techs or healthcare professionals – which have collected and stored data on individuals and/or businesses and have effective control over those data.

Data users: entities that receive and or process data shared by data providers on data subjects, as an input for providing a service to either the data subject or for the data user's own account.

Consent managers: a licensed fiduciary who is the intermediary between data subjects, data providers and data users and – as an advocate for data subjects – ensures that the agreed rules for data-sharing and processing are being followed.

Data generally comprise two classes: personal and non-personal data. Any data that are linked to a data subject's identity – ie they are personally identifiable – come under the category of personal data. From a data protection perspective, it is the data subject's personal data that are in the records (and under the control) of data providers, and these data are subject to the rights of data portability. These two classes of data – personal and non-personal data – can be further disaggregated into constituent components. The data categories are:

Personal data

a. *Non-shareable data* – such as passwords and biometric data – are unique to the market participant. As the name suggests, this category of data is not subject to sharing, although it can be required to access a platform, subject to the relevant cyber security standards. In many countries, biometric identity, card personal identification numbers (PINs) etc form this class of data and subscribe to technology standards prevalent in the country.

b. *Profile data* – such as name, date of birth, gender and address, or health-related data such as vaccination status – comprise characteristics that change infrequently. While service providers quite often require the provision of such data to access their platform, user consent is required for sharing this category of data (often through electronic know-your-customer (eKYC) or digital lockers).

c. *Generated data* – such as payments or borrowing transactions – are digital trails generated by the online activities of a consumer or SMEs. These data pertain to the activities of data subjects (consumers or SMEs) and consent is required for sharing this category of information. The service provider, in whose systems these data often reside, usually acts as the custodian for such data.

d. *Derived data* – such as credit scores – are developed by service providers through the mapping of generated data together with other attributes such as income, education status etc across users. In contrast to generated data, derived data consist of data drawn from the analysis of many data subjects (in the case of the big tech firms, millions of data subjects). That said, as such data could not be derived without the original generated data, the data subject has at least a partial claim to them, and some form of consent should be required for sharing data for which their generated data were an input.

Non-Personal data

a. *Anonymised data* – such as transaction data after removing personal identifiers – do not map to the identity of data subjects. These large data sets are crucial for scientific innovation (eg medical research or financial inclusion to name but a few). Typically, consent is not required for the sharing of these data, which need to meet anonymisation standards prevalent in the country to limit triangulation and ensure data anonymity.

b. *Public data* – such as GDP, Covid infection and death rates, retail sales or jobs-related data – that are regularly released by the government and research institutes. These data provide crucial input for the conduct of public policy. Consent is not required for assembling these data sets (although, when based on surveys, the input is likely to be anonymous), which are regulated by law and subject to open standards.

3. Consent-based data governance

There are two building blocks for an effective consent-based data-sharing system.

- A *data protection policy framework*. It is common practice that privacy (and data rights) is defined through domestic laws that recognise privacy (and data rights). The proposed framework recognises the rights of data subjects over the data they create – whether these data reside with them or not. Further, the data governance system asks for the consent of the data subjects prior to the processing and sharing of data, which is consistent with the provision of notice prior to the collection, sharing and processing of data.
- A *technological infrastructure that enables a user-friendly implementation of the data protection policy framework*. The technological infrastructure should be sector-agnostic to allow for cross-sectoral applications. To accommodate the many players involved such as financial and data services providers as well as both public and private sector institutions, the platforms upon which the infrastructure operates should be open, interoperable and non-discriminatory. At the same time, their design should ensure data security.

Given the need for granularity of consent, and given that data are to be spread among many users and providers, the quantity of data that need to be managed within the system is enormous. To achieve cost savings, data management must be digitally based and scalable across large numbers of users.

a. The conditions for data-sharing

When data are shared between data providers and data users, the data governance system should specify which data are requested for sharing, how long they will be retained by data users, and who will process them. In these areas, the system should meet the following five standards.

i) *purpose limitation*, ie ensure that the purpose for which data are being shared is described in clear and specific terms;

ii) *data minimisation*, ie share only as much data as are strictly necessary to achieve the stated purpose;

iii) *retention restriction*, ie ensure that data are not shared for longer than required to achieve the stated purpose;

iv) *use limitation*, ie ensure that data are used only for the purpose for which they were shared; and,

v) *operational resilience*, ie ensure that data are secure and the overall system is resilient to unauthorised access.

b. ORGANS principles[2]

State-of-the-art digital consent systems that replace the single act of giving consent to collect, process and share are built around the so-called ORGANS principles; namely, they are:

- ***O**pen:* they require the use of open, interoperable standards for the sharing of data. Since multiple players are involved in data-sharing – such as financial service providers, data services providers and data held by the government – the system must be open and interoperable.

- ***R**evocable:* they enable the revocation of consent once provided by the data subject; this includes the right to be forgotten, ie data subjects' data need to be deleted. Consent should be revocable once provided.

- ***G**ranular:* To obviate the need for provision of broad and sweeping ex ante consent, the granting of consent should be made more granular, specifying to whom data are provided, for how long and for what purpose. The system should allow for data subjects to provide consent in granular fashion just before data are shared. Granular consent requests – specifying which data are requested, how long they will be retained, and who will process them – satisfies purpose limitation, data minimisation, retention restriction and use limitation. Such granular consent can be granted on a recurring basis under similar conditions.

- ***A**uditable:* give data subjects the right to audit data-sharing transactions. Ease of audit requires machine-readable records of all consent provided by data subjects. Data subjects should have the right to audit data-sharing transactions ex post.

- ***N**otice:* recognise consent as a requirement for processing and sharing of data and require a notice of consent prior to collection, processing and sharing of data. The notice of consent sets forth the obligation to obtain the informed consent of the data subjects with the granular details described above prior to the collection, processing and sharing of data.

- ***S**ecure:* they impose data security obligations on data controllers. The consent artifact must subscribe to the highest standards of data providers and data users.

In Table 1, we provide the template that describes the proposed data governance system based on the above-mentioned conditions and principles. It includes both a data protection policy framework, whereby data rights and privacy are defined through laws that recognise rights and privacy on a national basis, and a technology infrastructure that enables a user-friendly sectoral implementation of the framework though software code.

[2]This section draws on NITI Aayog, "Data empowerment and protection architecture," August 2020.

Table 1: Granular template underpinning the proposed data governance system

	Data protection policy framework	Technology infrastructure
	The jurisdiction or one or more sectors within the jurisdiction has laws/policies that:	*In one or more sectors or cross-sectoral, standard interoperable technology structure is in place that:*
Data rights:	Recognises the rights of the data subjects over their data even if such data are under the control of data providers (controllers).	Enables the sharing of data between data providers (controllers) and data users with the electronic consent of data subjects.
Framework:	Governs the sharing of data.	Enables the sharing of data.
Open:	Requires the use of open interoperable standards for the sharing of data.	Enables the sharing of data that has been built using open, interoperable standards.
Interoperability of framework:	The same framework applies across sectors	The same framework applies across sectors
Notice and consent:	Recognises consent as a legitimate ground for processing and sharing of personal data and require notice prior to collection, processing or sharing of data.	Implements the principles of notice and consent in an electronic format.
Consent manager:	Allows for the establishment of consent intermediaries.	Implements a digital framework for consent intermediaries.
Granular:	Allows for consent to be provided in a granular fashion just before the data are shared.	Implements granular consent for data-sharing electronically.
Revocable:	Enables the revocation of consent once provided.	Implements the electronic revocation of consent for data-sharing.
Auditable:	Gives users the right to audit data-sharing transactions.	Enables the electronic audit of data-sharing transactions.
Data security:	Imposes data security obligations on data providers (controllers) and data users.	Builds data security into the design of the infrastructure.
Consumer adoption:	Implements measures to encourage consumer adoption such as simplification obligations, standardisation norms etc.	Encourages consumer adoption through, inter alia, inclusive user interfaces, transaction simplification for first-time digital users etc.

4. Data Empowerment Protection Architecture (DEPA) in India[3]

We next describe the evolving application of the consent-based data governance system in India and demonstrate its significant correspondence with the principles outlined above. India has utilised a financial technology stack in which a unified, multi-layered set of public sector digital platforms combine to provide substantial benefits to the population, from promoting financial inclusion and increasing efficiency to enhancing financial stability. A data governance system is the next critical layer in the India stack.

An effective data governance system encompassing notice and consent, purpose limitation, data minimisation, retention restriction and use limitation can only be implemented digitally. In India, the data protection policy framework that defines how personal data can be collected and processed has a technological framework as its counterpart: the principles of DEPA are implemented in software codes.

a. Application of DEPA to the financial sector

While the overall data protection policy framework has yet to be enacted in law, it has been adapted to – and is operational in – the financial sector through the Account Aggregator system, which went live in September 2021 (Graph 1).[4]

Data sharing within DEPA works as follows. Before a data subject initiates a data transfer, the subject needs to enrol with an Account Aggregator, or consent manager (CM), and in so doing provides a list of approved data providers/controllers to the consent manager (1). When a data subject seeks service from a data user (2) – which, for instance, can comprise parties such as lenders, insurance companies and personal finance managers – the data user initiates a data transfer request (3), which is submitted to the CM. The data user chooses a template from a suite of templates designed for this – specifying the purpose of the data transfer, the specific data that are needed to satisfy that purpose and the duration for which they will be retained – and picks the data request format that meets the requirements of the request. While these templates cover a broad range of uses for which data may be requested, at the same time they ensure that only the minimum amount of data needed for the purpose at hand is requested, thus meeting the principles of notice, consent, purpose limitation and data minimisation.

Only after the data subject has provided the consent for sharing data (4) does the CM submit this request to the data providers (5). The data providers, in turn, can include financial information providers, tax platforms and insurance providers, among others. After verifying the request, the data provider transfers the data through an end-to-end encrypted flow to the consent manager (6), who shares the data with the data user.

In this series of transactions, the CM is aware of the identity of both the data users and data providers, but blind to the content of the data that the CM is transferring. Data users, on the other hand, are aware of the content of the data but blind to the identity of the data provider. Similarly, data providers are aware of the content of the data but blind to the identity of the data user. Through the consent manager, data flows are separated from consent flows, thereby ensuring the efficient transfer of data while respecting privacy concerns.

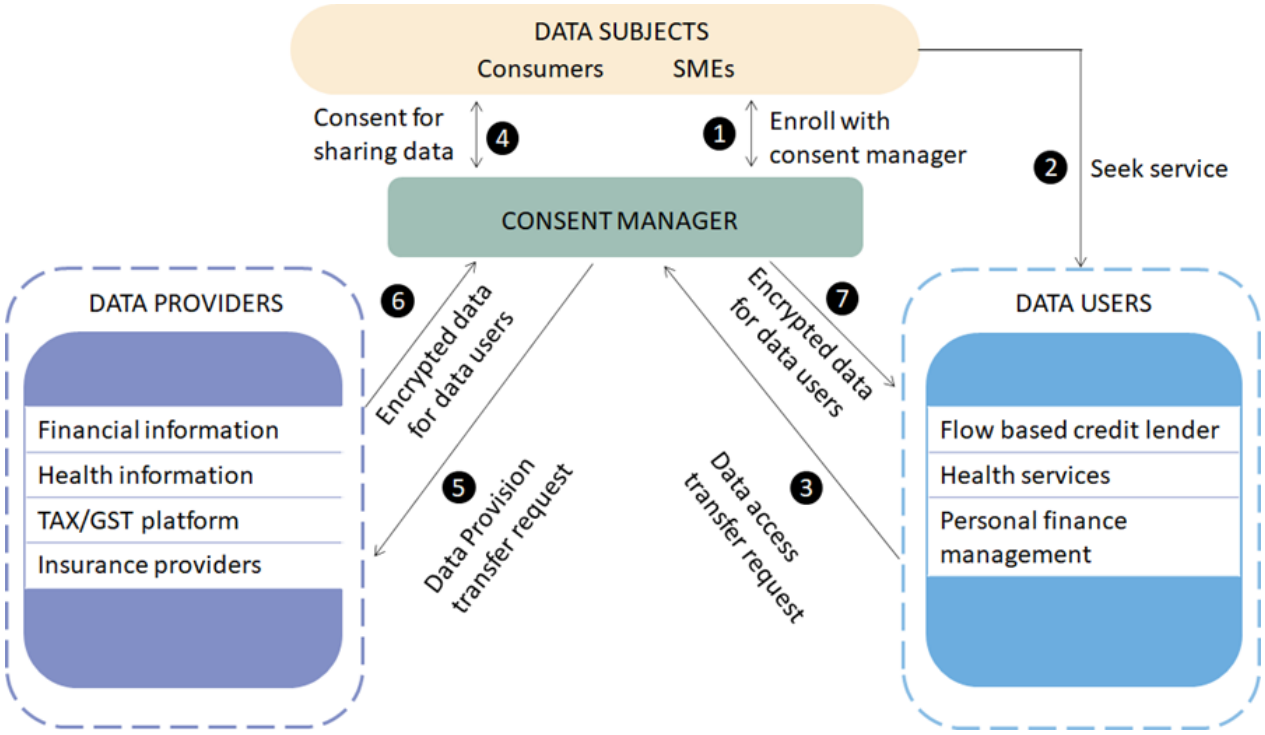
[3]The discussion in this section draws on the previously cited NITI Aayog (2020), and iSPIRT Foundation, Data Empowerment: A Techno-Legal Approach, 2021.

[4]See Reserve Bank of India, "Master direction – non-banking financial company – Account Aggregator directions, 2016" (Updated as on 5 October 2021). The Data Protection Bill, which covers a much broader mandate of privacy protections, is expected to be enacted into law in 2022.

This structure, while vastly improved over most earlier consent systems, does not ensure that the data user is using the information only for the purpose for which data were shared or keeping them only for the period initially agreed. In other words, once the data are shared with the data user, there is no feature in this architecture that can assure the data subject that the principle of use limitation will be satisfied. Thus, the next advance in this architecture will be the addition of a confidential clean room environment in which the data user can process the data and extract the results of the analysis but not the personally identifiable information data themselves. Because data never leave the execution environment, this architecture, when it becomes available, will provide a high degree of assurance regarding use limitation and increase incentives to share data.

When we benchmark the data governance architecture in India against the granular template outlined earlier in Table 1, we find that nearly all of the elements necessary for optimal data governance are in place. The only caveat is that the data policy framework and interoperable technology infrastructure has to date only been developed for sectors in the financial services industry.

Graph 1: The data sharing system as applied by DEPA to the financial sector



Source: Authors' elaboration.

b. Early results[5]

Early results give some indication of the ability of the system to be scaled up. As of 14 April 2022, nine banks were fully operational as financial information providers (FIPs). These nine banks have a combined total of 215 million individual savings, term deposits plus sole proprietor current accounts. On the other side, there are 35 financial information users (FIUs), predominantly non-banks, banks and a handful of registered investment advisors that are fully operational, or live. There are 10 licensed consent managers known as Account Aggregators, of which 50% are fully operational.

[5]All data are sourced from Sahamati, https://sahamati.org.in/.

A larger number of entities across various financial sectors are currently engaged at some level in the data sharing ecosystem (either live, testing, in the tech development stage or evaluating). When participation is defined this broadly, 41 banks are engaged as FIPs and FIUs, while 56 non-banks are engaged are FIUs. In total, 56 institutions are engaged as FIPs and 133 are engaged as FIUs.

In terms of the system's actual usage, 230,000 consent requests from the FIUs were processed over the initial 30 weeks, thus averaging around 1,000 consent requests daily. In about 90–95% of cases, the FIPs have successfully addressed consent requests with an average response time in seconds. The system is market-driven, with participants renumerated for the costs they incur. Evidence on costs, provided in BIS Papers No. 124, indicates that irrespective of the suite of services offered, under current circumstances, the marginal cost of a data pull to consumers is modest.

5. Conclusion

In this paper, we have proposed a data governance system that restores control to data subjects with regard to the collection, processing and sharing of their data. This framework – which replaces broad and sweeping consent with granular consent provided on digital basis – has been influenced by the privacy laws prevalent in many jurisdictions. Given the granularity and amount of data spread over numerous data subjects and data controllers, only a digital system can be secure and operate at low transaction costs. Thus, technological protocols for notice and consent, purpose limitation, data minimisation, retention restriction and use limitation play a key role in operationalising the data governance framework. India's Data Empowerment Protection Architecture (DEPA), which went live in the financial sector in September 2021, is an example of a data governance system following such a template. Evidence from the early experience of DEPA suggests that such a consent-based system can operate at scale with low transaction costs. ■

About the authors

Siddharth Tiwari became Chief Representative of the BIS Office for Asia and the Pacific in November 2018. Before joining the BIS, Mr Tiwari served as Executive Secretary to the G20 Eminent Persons Group on Global Financial Governance from 2017 to 2018. From 1985 to 2017, he was at the IMF, where he occupied top-level positions and shepherded the institution's work on strategy, policy and lending operations, as well as on administrative matters. He also has significant experience in international policymaking and cooperation, including on issues of great relevance to the Asia-Pacific region. He earned his master's and doctorate degrees in economics from the University of Chicago.

Sharad Sharma is part of the Governing Council and Co-Founder of iSPIRT (Indian Software Product Industry RoundTable) Foundation. Sharad has three decades of experience in the internet, enterprise software, and digital infrastructure markets and is a prominent voice in India's technology ecosystem. Previously, he was the CEO of Yahoo! India R&D. Sharad also co-founded Teltier Technologies Inc., a wireless infrastructure startup, now part of CISCO, and is an active technology angel investor with over two dozen investments. Sharad has held several senior R&D executive positions with leading technology companies, including VERITAS Software, Symantec, Lucent Technologies, and AT&T. Sharad is a Member of SEBI's Financial and Regulatory Technology Committee. He has also served on RBI's UK Sinha MSME Committee and National Digital Payments Committee. He did his Electrical Engineering at Delhi College of Engineering.

Siddharth Shetty is a Fellow at iSPIRT Foundation, a non-profit technology think tank, where he works on India Stack, a set of Open APIs that allows governments, businesses, startups and developers to utilise a unique digital Infrastructure to solve India's hard problems in financial inclusion. As part of India Stack, his primary focus is on empowering every Indian with control of their financial, health, telecom, skills, and education data through the Data Empowerment and Protection Architecture (DEPA). Besides for DEPA, he's also working on the technology for the Public Credit Registry by Reserve Bank of India, the digital infrastructure that would enable safe operations of millions of drones in Indian airspace; and the National Health Stack that would lower costs, improve access, and improve quality of healthcare for hundreds of millions of Indians.

Frank Packer is a Regional Adviser at the BIS Representative Office for Asia and the Pacific. Before moving to the BIS Representative Office for Asia and the Pacific, Frank Packer was head of Financial Markets in the BIS's Monetary and Economic Department, and editor of the BIS Quarterly Review. Prior to assuming the current position, he was line manager in the Asian Office for six years, most recently as Head of Economics and Financial Markets. Earlier in his career, he worked for the Federal Reserve Bank of New York and Nikko Citigroup in Tokyo. He received his PhD from Columbia University, an MBA from the University of Chicago and a BA from Harvard. His research interests currently include issues related to digital financial infrastructure, green and transition finance, and central banking.

SUERF Publications

Find more **SUERF Policy Notes** and **Policy Briefs** at www.suerf.org/policynotes

SUERF
The European Money and Finance Forum

SUERF is a network association of central bankers and regulators, academics, and practitioners in the financial sector. The focus of the association is on the analysis, discussion and understanding of financial markets and institutions, the monetary economy, the conduct of regulation, supervision and monetary policy. SUERF's events and publications provide a unique European network for the analysis and discussion of these and related issues.

SUERF Policy Notes focus on current financial, monetary or economic issues, designed for policy makers and financial practitioners, authored by renowned experts.

The views expressed are those of the author(s) and not necessarily those of the institution(s) the author(s) is/are affiliated with.

SUERF Secretariat
c/o OeNB
Otto-Wagner-Platz 3
A-1090 Vienna, Austria
Phone: +43-1-40420-7206
www.suerf.org • suerf@oenb.at