BIS

# BIS Bulletin

No 87

## Next generation correspondent banking

Rodney Garratt, Priscilla Koo Wilkens and Hyun Song Shin

30 May 2024

BIS Bulletins are written by staff members of the Bank for International Settlements, and from time to time by other economists, and are published by the Bank. The papers are on subjects of topical interest and are technical in character. The views expressed in them are those of their authors and not necessarily the views of the BIS. The authors are grateful to Iñaki Aldasoro, Claudio Borio, Agustin Carstens, Sonja Davidovic, Maha El Dimachki, Anneke Kosse, Thomas Lammer, Tara Rice, Beju Shah, Takeshi Shirakami and Peter Wierts for comments and suggestions, to Ilaria Mattei for research support and to Louisa Wagner for administrative support.

The editor of the BIS Bulletin series is Hyun Song Shin.

This publication is available on the BIS website (www.bis.org).



ISSN: 2708-0420 (online)
ISBN: 978-92-9259-761-00 (online)

Rodney Garratt
*Rodney.Garratt@bis.org*

Priscilla Koo Wilkens
*Priscilla.KooWilkens@bis.org*

Hyun Song Shin
*HyunSong.Shin@bis.org*

# Next generation correspondent banking

*Key takeaways*

- *Existing correspondent banking processes have struggled to adapt to new regulatory and supervisory requirements, posing questions on the future of the correspondent banking model.*
- *The tokenisation of correspondent banking, as embodied in Project Agorá (BIS (2024b)), could unlock streamlined pre-screening and atomic settlement, and pave the way for superior customer verification and anti-money laundering (AML) procedures.*
- *Tokenisation could substantially reduce duplication and miscoordination, thereby revitalising cross-border payments by fostering a robust network of correspondents and corridors.*

Correspondent banks played a pivotal and historical role in the expansion of cross-border payment activity by enabling transaction settlement, cheque clearance and associated foreign exchange operations. However, the correspondent banking system has seen a reduction in the number of active correspondents and corridors in recent years (Rice et al (2020)). Is the observed decline evidence that correspondent banking is a flawed model? Or is it simply the case that a technological overhaul is needed to help this age-old system meet the more stringent user and regulatory requirements of the 21st century? This Bulletin explores how tokenisation through the application of the BIS's unified ledger concept (BIS (2023)) could enhance the functioning of correspondent banking by harnessing new technology to address key pain points in cross-border payment chains. The recently announced Project Agorá (BIS (2024b)) is a concrete initiative in this direction.

## How correspondent banking works

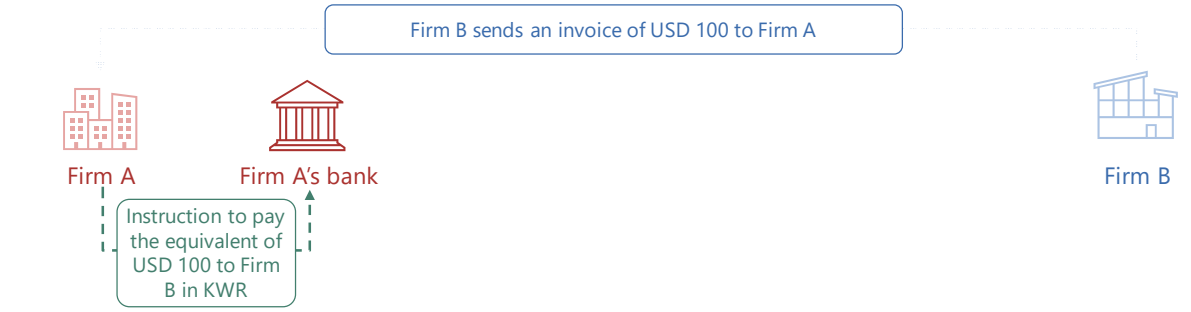
Speaking of "sending a payment" conjures up the image that a payment involves the transfer of an object. However, when a customer of one bank "pays" a customer of another bank, no physical object changes hands. Rather a series of updates are made on the accounts maintained by intermediaries. In a domestic payment, this is accomplished relatively easily because the central bank settles the payment between intermediaries. The payee's bank credits the payee's account because the bank receives equal value in central bank money through an update of the central bank ledger. However, in a cross-border context there are often different sovereign currencies involved and hence there is no single common settlement asset that can be used throughout the transaction chain. In such cases, another approach is needed.

Graph 1 depicts an example to fix ideas. Firm A is a manufacturing firm located in the United States who has received the latest batch of intermediate inputs from its supplier, Firm B, located in Korea. The invoiced amount is USD 100, and Firm A has instructed its bank to proceed with the payment (Graph 1.A). If Firm B had a US bank account, then the payer's bank would debit the payer's account, while the payee's bank would credit the payee's account, with the central bank settling the payment between the two banks by transferring reserve balances. However, what happens if Firm B does not have a US bank account and instead needs to receive the Korean won equivalent of USD 100 in its Korean bank?

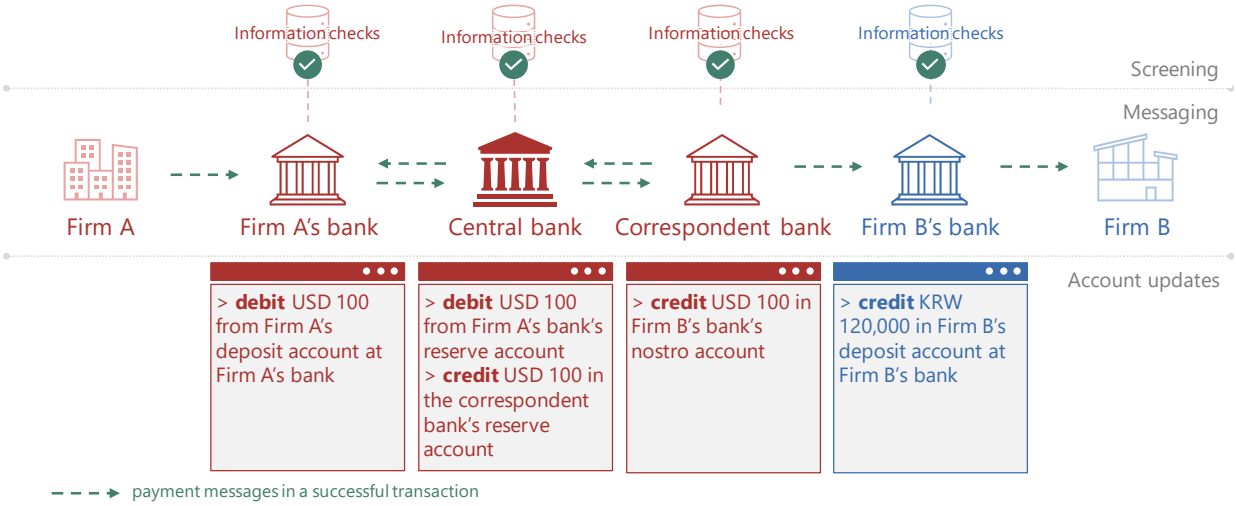
## Mechanics of correspondent banking

Graph 1

A. Making a cross-border payment entails bridging the gap between the payer's and payee's banks




B. Correspondent banking bridges the gap through messages and balance sheet updates




Source: Authors' elaborations

Graph 1.B illustrates how a correspondent bank can bridge the gap between Firm A and Firm B. Firm A's bank makes a payment to the correspondent bank for the invoiced amount of USD 100. Since this is a purely domestic payment within the United States, the payment follows standard procedures and is settled on the balance sheet of the Federal Reserve. After this domestic interbank payment, the correspondent bank sees an increase of USD 100 in its reserve balance at the Federal Reserve and then credits the account of the Korean bank by USD 100. Finally, the Korean bank verifies that it has received USD 100 in its account at its US correspondent bank and then credits Firm B's account in Korea by the equivalent amount in won. Notice that this cross-border payment does not require settlement across countries. Settlement occurs only in the United States.

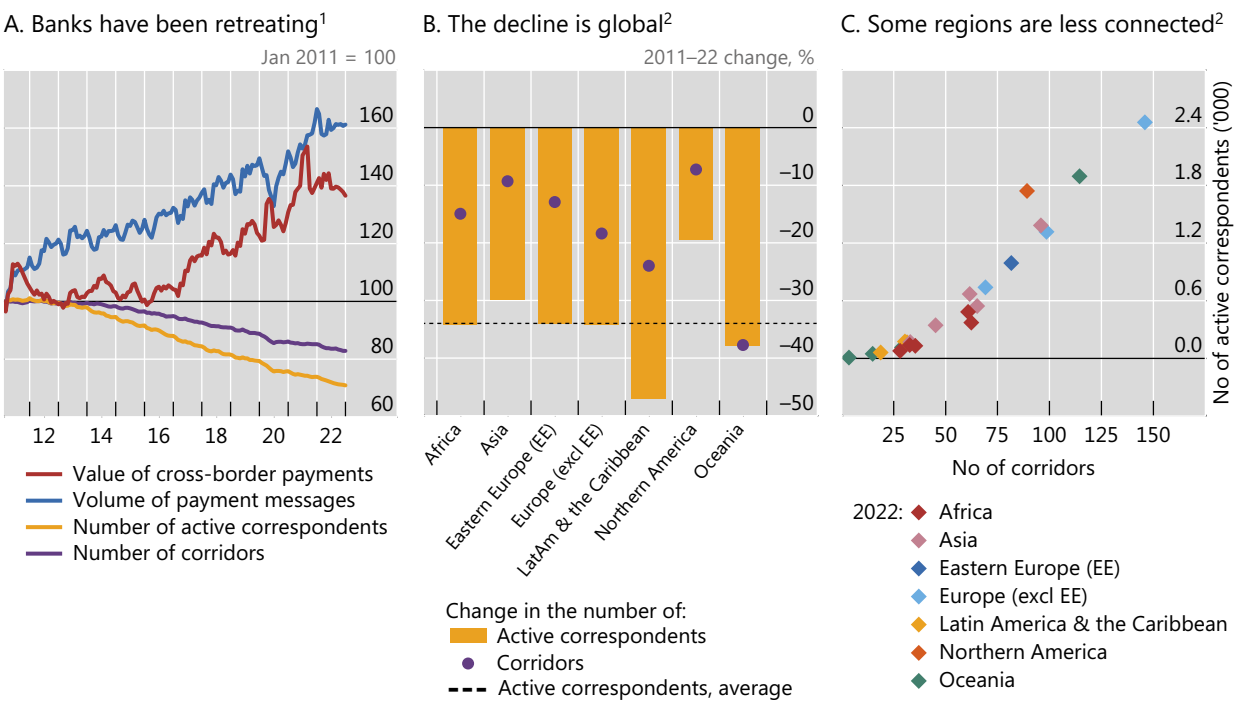
Many messages are required to coordinate the balance sheet updates along the payment chain (middle of Graph 1.B). Historically, correspondent banking arose from informal exchange built on the chain of trust relationships between merchants who acted both as trade counterparties as well as financiers (hence the term "merchant banker"). Financial instruments such as bills of exchange (and later cheques) met the need for working capital as well as serving as payment instruments (Schnabel and Shin (2004), James and Weiman (2010)). The system necessitated the physical transfer and manual updating of documents. Digital record keeping and messaging platforms such as SWIFT have increased speed and efficiency, but our current system still bears the hallmarks of its historical origin in the way that the system works through chains of bilateral relationships.

Meanwhile, important integrity cornerstones such as know your customer (KYC) and AML compliance requirements have been layered on top of the chained process with its origins in more informal, trust-based modes of exchange. In today's more rigorous environment, integrity safeguards are key. In addition to the requirement that each message be complete, consistent and error free, the payment amounts must comply with restrictions, the recipients cannot be on sanctions lists, and each bank must perform AML and countering the financing of terrorism (CFT) checks (top of Graph 1.B). The burden of performing these checks is placed on individual banks, who follow a sequential process that takes time and involves significant duplication of effort. When problems emerge, steps already taken need to be unwound. In fact, failures to complete cross-border payments are frequent and expensive, impacting not only direct costs of unwound transactions but also the costs of uncompleted business (Borchert et al (2023)).

## Trends in correspondent banking

As commerce has become more global and payment chains more complex, trust relationships have given way to stringent controls to ensure the integrity of the payment system. Adherence to these controls requires capital investments in technological infrastructure, personnel training and meticulous regulatory reporting systems. It also entails absorbing the risk of penalties from inadvertent underreporting of criminal activity. Anecdotal evidence suggests banks are shifting their business models away from cross-border payment services along some corridors due to the unfavourable risk-return trade-off.[1] The number of active correspondents and corridors decreased between 2011 and 2022 (Graph 2.A) and the trend is

Correspondent banking relationships have been declining — Graph 2




[1] Three-month moving averages. [2] Averages across countries in each region. Correspondent banks that are active in several corridors are counted several times. Grouping of countries by regions is according to the United Nations Statistics Division; for further details see unstats.un.org/unsd/methodology/m49/.

Sources: Rice et al (2020); CPMI (2023); authors' calculations.

[1] Compliance risks are cited by banks as the primary reason for cutting correspondent ties, especially in countries with weaker governance (Everington (2023)).

global (Graph 2.B). Emerging markets are particularly affected by the shrinking of correspondent banking corridors (Graph 2.C), and increased costs have weighed on real economic activity (IFC (2017), Adrian et al (2023), Rice et al (2020)).[2] While incremental approaches to improving cross-border payments can mitigate some inefficiencies, they have not been able to address the key issues in a wholly satisfactory way. Instead, a new approach is needed. This is the promise of the unified ledger approach to tokenisation in the recently launched Project Agorá (BIS (2024b)).

## Next generation correspondent banking

The new financial market infrastructure (FMI) envisaged in Project Agorá is built around tokenised deposits and tokenised central bank money and is patterned after the unified ledger concept put forth in BIS (2023). The roles of institutions (commercial banks and central banks) and the roles of money (commercial bank deposits and central bank money) would remain unchanged from today, but the platform would harness three new capabilities: (1) combining payment messaging and account updates as a single operation; (2) executing the chain of payments atomically rather than as a series of sequential updates to separate ledgers; and (3) drawing on privacy-preserving platform resources for KYC/AML compliance. This new FMI would guarantee atomic settlement through synchronous execution of all commands and streamline pre- and post-trade workflows involved in cross-border payments, while ensuring "integrity by design" (FATF (2024)).

Tokenisation refers to the digital representation of claims on a shared programmable platform that incorporates the rules and logic governing transfers as well as the traditional database information underpinning current systems (Aldasoro et al (2023)). Tokenisation enables the joint execution of all steps in end-to-end payments, eliminating the need for sequential processing and the risk of reversals. On a shared platform, actions can be taken in parallel and made contingent on the completion of all required information checks and screening requirements (Graph 3.B). Any failures can be instantaneously communicated to all those involved before settlement happens.

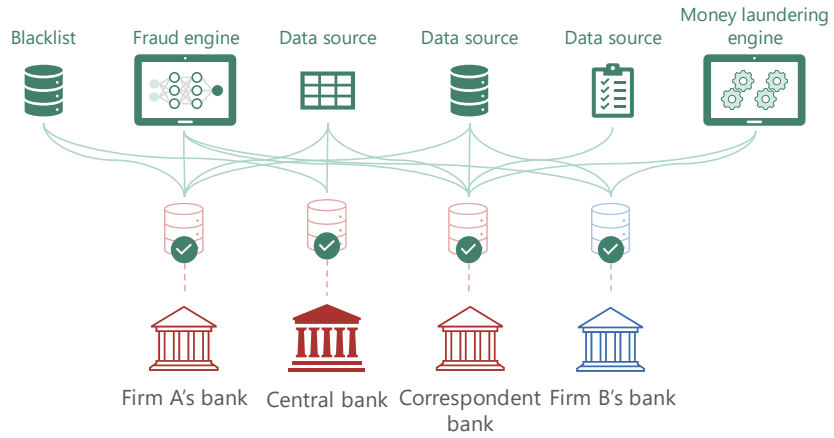
Pre-screening capabilities have already been developed in Project Mandala (BIS (2024c)). This proof of concept demonstrates the feasibility of encoding jurisdiction-specific policy and regulatory requirements into a common protocol for cross-border use cases. In the next generation correspondent banking platform envisaged in Project Agorá, the efficiency gains will also come from embedding pre-screening confirmations into the end-to-end payment transaction.

In addition, collaborative effort and privacy-preserving techniques for data-sharing bring increased scope for using machine learning and artificial intelligence to improve screening and AML compliance. A common problem with screening for sanctioned individuals is false positives, which arises when an individual has the same name or other identifying information as a sanctioned person. Supervised AI can be used to recognise when these errors arise by relating the transaction to past transactional activity involving the individual with the flagged name. Transaction surveillance and monitoring can also be done more efficiently and effectively by sharing data confidentially and doing collaborative monitoring. Enhanced AML procedures for detecting suspicious networks and flows of funds can be achieved using two different architectures. In one option, all transactional data are shared into a centralised, privacy- and security-enhanced data infrastructure. Another option is a decentralised privacy and security approach using federated learning, where models are trained on local data, and only insights are shared rather than the underlying data. The latter approach, which supports cross-border collaboration while addressing data sovereignty issues, is explored in Project Aurora (BIS (2024a)).

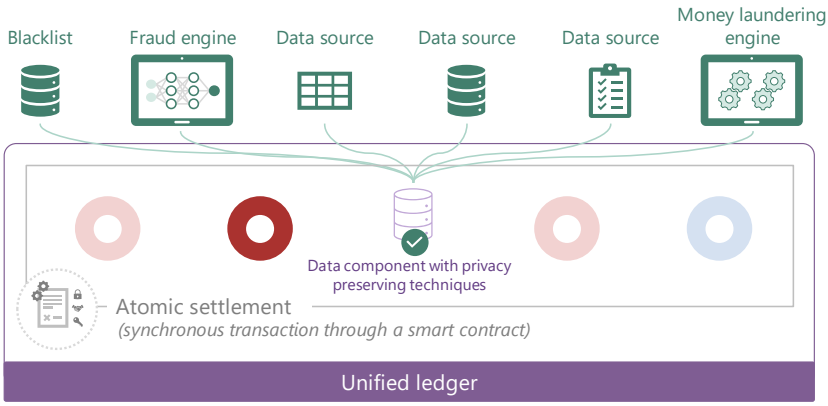
[2] The policy issues related to these developments are laid out in a report of the BIS Committee on Payments and Market Infrastructures (CPMI (2016)).

KYC/AML in next generation correspondent banking Graph 3

A. Silos lead to duplication of effort in AML screening




B. Collaborative effort and data-sharing




Source: Authors' elaborations

## The road ahead

Next generation correspondent banking harnesses the promise of tokenisation to create a more powerful and economically viable correspondent banking system. Such a system will not only make existing correspondent banking corridors more efficient, by reducing costs and improving services for customers, but also make previously closed corridors economically viable once again. For low-income countries, next generation correspondent banking will open the door to greater participation in global trade and remittance channels.

Banks are individually responsible for complying with sanctions, implementing KYC procedures and ensuring AML/CFT compliance. Adopting the new system will entail overcoming collective action problems on the part of current incumbents. Progress may require some accommodation in the form of rule changes. However, there can be no compromise on the integrity of global payments, and so turning back the clock is not an option.

Project Agorá (BIS (2024a) is a promising step towards next generation correspondent banking. Seven central banks will settle payments between banks and correspondents in their jurisdictions.[3] The

[3] The participating central banks are the Bank of France (representing the Eurosystem), Bank of Japan, Bank of Korea, Bank of Mexico, Swiss National Bank, Bank of England and the Federal Reserve Bank of New York. The project is managed by the BIS as a public sector convener and the International Institute of Finance as a private sector convener. An open call for participation will select private sector entities (BIS (2024b)).

development of a new payment infrastructure like Agorá should ensure potential gains are not lost due to fragmentation (Economist (2024)). Next generation correspondent banking should be inclusive, accessible and capture beneficial network effects. This can be accomplished via policies on access to the infrastructure or via interoperability with other infrastructures, as advocated in the Finternet vision (Carstens and Nilekani (2024)). Corridors that exist outside the included currencies could connect to the system to form a complete payments network.

## References


Adrian T, R Garratt, D He and T Mancini-Griffoli (2023): "Trust bridges and money flows", *BIS Working Papers*, no 1112, July.

Aldasoro I, S Doerr, S Gambacorta, R Garratt and P Koo Wilkens (2023): "The tokenisation continuum", *BIS Bulletin*, no 72, April.

Bank for International Settlements (BIS) (2023): "Blueprint for the future monetary system: improving the old, enabling the new", *Annual Economic Report 2023*, Chapter III, June, pp 85–118.

_____ (2024a): "Project Aurora: the power of data, technology and collaboration to combat money laundering", March.

_____ (2024b): "Project Agorá: central banks and banking sector embark on major project to explore tokenisation of cross-border payments", April.

_____ (2024c): "Project Mandala: shaping the future of cross-border payments compliance", May.

Borchert, L, R De Haas, K Kirschenmann and A Schultz (2023): "Broken relationships: de-risking by correspondent banks and international trade", *EBRD Working Papers*, no 285, December.

Carstens, A and N Nilekani (2024): "Finternet: the financial system for the future", *BIS Working Papers*, no 1178, April.

Committee on Payments and Market Infrastructures (CPMI) (2016): "Correspondent banking", *CPMI Papers*, no 147, July.

_____ (2023): CPMI quantitative review of correspondent banking data, May.

Economist, The (2024), "National payment systems are proliferating", *Worlds Apart*, Special Report, 11 May.

Everington, J (2023): "Solving correspondent banking's woes", *The Banker*, September.

Financial Action Task Force (FATF) (2024): "Declaration of the Ministers of the Financial Action Task Force", April.

International Finance Corporation (IFC) (2017): "De-risking and other challenges in the emerging market financial sector: findings from IFC's survey on correspondent banking", *IFC Insights*, September.

James, J and D Weiman (2010): "From drafts to checks: the evolution of correspondent banking networks and the formation of the modern U.S. payments system, 1850–1914", *Journal of Money, Credit and Banking*, vol 42, no 2–3, pp 237–265.

Rice T, G von Peter and C Boar (2020): "On the global retreat of correspondent banks", *BIS Quarterly Review*, March, pp 37–52.

Schnabel I and H S Shin (2004): "Liquidity and Contagion: The Crisis of 1763", *Journal of the European Economic Association,* vol 2, no 6, pp 929–68.

## Previous issues in this series

| | | |
|---|---|---|
| No 86<br>23 April 2024 | Survey evidence on gen AI and households: job prospects amid trust concerns | Iñaki Aldasoro, Olivier Armantier, Sebastian Doerr, Leonardo Gambacorta and Tommaso Oliviero |
| No 85<br>28 March 2024 | From financial inclusion to financial health | Carlos Cantú, Jon Frost, Tirupam Goel and Jermy Prenio |
| No 84<br>23 January 2024 | Artificial intelligence in central banking | Douglas Araujo, Sebastian Doerr, Leonardo Gambacorta and Bruno Tissot |
| No 83<br>04 January 2024 | Testing the cognitive limits of large language models | Fernando Perez-Cruz and Hyun Song Shin |
| No 82<br>20 December 2023 | The contribution of monetary policy to disinflation | Pongpitch Amatyakul, Fiorella De Fiore, Marco Lombardi, Benoit Mojon and Daniel Rees |
| No 81<br>13 December 2023 | Interest rate risk of non-financial firms: who hedges and does it help? | Ryan Banerjee, Julián Caballero, Enisse Kharroubi, Renée Spigt and Egon Zakrajšek |
| No 80<br>23 November 2023 | Monetary policy, financial conditions and real activity: is this time different? | Fernando Avalos, Deniz Igan, Cristina Manea and Richhild Moessner |
| No 79<br>2 November 2023 | Lessons from recent experiences on exchange rates, capital flows and financial conditions in emerging market economies | Pietro Patelli, Jimmy Shek and Ilhyock Shim |
| No 78<br>3 October 2023 | Mapping the realignment of global value chains | Han Qiu, Hyun Song Shin and Leanne Si Ying Zhang |
| No 77<br>13 September 2023 | Margins and liquidity in European energy markets in 2022 | Fernando Avalos, Wenqian Huang and Kevin Tracol |
| No 76<br>7 September 2023 | The oracle problem and the future of DeFi | Chanelle Duley, Leonardo Gambacorta, Rodney Garratt and Priscilla Koo Wilkens |
| No 75<br>19 May 2023 | Disinflation milestones | Benoit Mojon, Gabriela Nodari and Stefano Siviero |
| No 74<br>13 April 2023 | The changing nexus between commodity prices and the dollar: causes and implications | Boris Hofmann, Deniz Igan and Daniel Rees |
| No 73<br>11 April 2023 | Stablecoins versus tokenised deposits: implications for the singleness of money | Rodney Garratt and Hyun Song Shin |

All issues are available on our website www.bis.org.