BIS Innovation Hub

Project Atlas

Mapping the world of decentralised finance

October 2023

DeNederlandscheBank
EUROSYSTEEM

DEUTSCHE BUNDESBANK
EUROSYSTEM

Content

Executive summary

Project Atlas creates a data platform that sheds light on the macroeconomic relevance of cryptoasset markets and decentralised finance (DeFi). Together with the project partners within the Eurosystem – the Deutsche Bundesbank and De Nederlandsche Bank – a first proof of concept of Project Atlas was developed focusing on international flows of cryptoassets.

Cryptoassets and DeFi applications are part of an emerging financial ecosystem that spans the globe. While introducing new technologies, these markets often lack transparency and potentially present risks to financial stability. The collapse of some stablecoins and DeFi platforms has highlighted the difficulty of making such risk assessments today. Although blockchain transactions are theoretically transparent, reliable information on macro-financial implications is hard to obtain.

Project Atlas provides data tailored to the needs of central banks and financial regulators. It fuses data gathered from crypto exchanges (off-chain data) with data from public blockchains (on-chain data) gathered from nodes. By connecting various sources, Atlas allows for data vetting, giving users tools to evaluate these markets' economic significance more accurately.

As part of a first proof of concept, Project Atlas derives cryptoasset flows across geographical locations. The approach uses transactions attributed to crypto exchanges in the Bitcoin network, along with the location of those exchanges, as a proxy for cross-border capital flows. The country location is not always discernible for crypto exchanges, and attribution data are naturally incomplete and possibly not perfectly accurate. Therefore, the flows should be regarded as a lower-bound estimate of the actual size.

The initial findings indicate that, although relatively small compared with total on-chain network traffic, identified flows between crypto exchanges are significant and substantial economically. Attributing geographical areas to exchanges (where possible) lays out the structure of cross-border flows. Thus, Project Atlas provides a starting point for structural analysis across jurisdictions.

More broadly, there is a need for central banks and financial regulators to gain first-hand knowledge of cryptoasset and DeFi markets, and there is a dearth of reliable and tailored data for such purposes. Policymakers must understand the underlying data that feed into aggregate indicators to make well informed decisions. Available aggregate statistics provided by market actors or data providers often leave open how data are generated and what the underlying assumptions are. Access to granular data supports assessment of data reliability and enables solid analysis.

Because Project Atlas relies on in-house development of the platform and broader infrastructure, the knowledge and developed code can be openly shared with the central banking community. At the same time, Atlas enhances technical and analytical capabilities. Atlas can evolve into an insightful public good as the data platform and outputs will be openly available to central banks.

1 Introduction

Introduction

Cryptoassets and decentralised finance (DeFi) applications are part of an emerging financial ecosystem that spans the globe.[1] Since Bitcoin debuted in 2009, the ecosystem has seen astounding growth. Today, there are thousands of cryptoassets, with a collective market capitalisation of more than $1 trillion. With its numerous actors, protocols and networks, crypto and DeFi form a convoluted ecosystem that is difficult to analyse.

While introducing new technologies, crypto markets often lack transparency and present risks. Unique risks emerge due to the decentralised setup and emerging new intermediaries such as crypto exchanges and stablecoin issuers. Recently, the exploit of Curve Finance pools exposed vulnerabilities across DeFi projects, creating contagion risks.[2] The crash of the Terra (Luna) protocol's algorithmic stablecoin in a downward spiral and the bankruptcy of centralised crypto exchange FTX also highlight the pitfalls of unregulated markets.

Regulators are becoming increasingly apprehensive about the scale of crypto markets and their integration with traditional finance. Despite banks' minimal exposure to crypto markets to date (Auer et al (2022)), financial stability concerns are intensifying (FSB (2022)).

The question then becomes: do central banks have the data and tools to effectively measure the evolution of crypto and DeFi markets? Crypto markets are highly dynamic, and while policymakers need reliable statistics on these markets, they are often elusive. Surprisingly, even calculations of basic indicators, such as on-chain transaction volumes, frequently differ across sources due to different methodologies. This calls for further investigation and the ability to replicate statistics.

In theory, data on blockchain transactions are openly available and transparent. However, they are vast and complex, making them challenging to organise and analyse and to derive economic meaning from them. Cryptoasset analytics requires taking into consideration the ever-changing nature of the ecosystem, which affects both the gathering of data and the evaluation of meaningful indicators.

Blockchain data are only one piece of the puzzle. The DeFi ecosystem now includes centralised intermediaries, such as crypto exchanges and custodians. Contrary to DeFi's intent to provide financial services without centralised entities, concentrations of power and centralised governance erode the essence of decentralisation in this space (Aramonte et al (2021)).

The emergence of new actors makes it difficult to map the ecosystem using data. Oftentimes, unsupervised providers rather than regulated banks offer financial applications in this realm. Thus, reported data lack standardisation and adequate

[1] See Auer et al (2023) for a definition and description of the underlying technology.

[2] See eg cointelegraph.com/news/curve-vyper-exploit-whole-story-so-far.

checks. Non-adherence to existing regulations, the difficulty of enforcing regulatory requirements and operation of market actors outside their scope (Deutsche Bundesbank (2021)) compound the risk of manipulated or distorted data. Market actors have incentives to exaggerate self-reported data in order to appear larger or safer, thereby attracting more business. As much as 70% of trading volumes reported by unregulated exchanges is estimated to constitute wash trading (Cong et al (2023)).

Crypto and DeFi markets leave open considerable data gaps. For example, information on macro-financial implications, such as cross-border capital flows, is hard to obtain. Project Atlas seeks to answer the question: how can central banks harness technologies to gain insights into DeFi markets and gather more reliable and comprehensive statistics on crypto ecosystems?

Project Atlas aims to close some of the existing data gaps by developing a specialised data platform encompassing various data sources to illuminate the macroeconomic significance of crypto and DeFi. The proof of concept (PoC) provides insights into cross-border flows. Tailored to the needs of central banks and financial regulators, it fuses data gathered from crypto exchanges (off-chain data) with public blockchain data (on-chain data) gathered via operating a node. Off-chain data include any data not extracted from the ledgers, such as trades that crypto exchanges settle internally. The benefits of connecting off- and on-chain data lie in creating novel statistics and the vetting of existing ones.

Atlas seeks to give policymakers, analysts and researchers access to granular data and custom indicators. The platform operates in a cloud sandbox and can integrate new data sources over time. Policymakers will be able to evaluate developments via dashboards, while analysts and researchers can delve into specific research questions using the analytics platform.

Atlas integrates diverse data sources and builds up robust technical and analytical capabilities to support the central bank community. While the data are often openly available, central banks may not always be able to process them effectively due to their complexity and high computational demands. Atlas serves as a public good and will provide central banks worldwide with the opportunity to harness its insights and resources. Because Atlas relies on in-house development of the platform and broader infrastructure, the knowledge and developed code can be openly shared with the central banking community.

As part of a first PoC, Project Atlas is focusing on bitcoin flows between crypto exchanges across geographical locations.[3] Although relatively small compared with total network traffic, the data show that flows between crypto exchanges are non-negligible. Attributing geographical locations to exchanges (where possible) lays out the structure of cross-border flows and provides a starting point for evaluating the relative economic significance across jurisdictions. For example, the data will allow flows to be analysed structurally, and the influence of price shocks, financial market

[3] The data described here are based on the location of crypto exchanges for which flows are identified, not on the users' location of these exchanges.

developments and country characteristics on crypto flows to be investigated. The data also show the size of holdings in different jurisdictions over time.

The next section provides background on Project Atlas's motivation as well as a discussion of crypto and DeFi data sources. The annex gives an introduction to crypto and DeFi. Section 3 provides background on the technical architecture; readers interested in the concrete approach of the PoC and visualisations may skip this section. Section 4 showcases the results of the first application of Atlas to cross-border flows. The dashboards Atlas uses to visualise these flows are described in Section 5. Section 6 concludes.

2 The need for tailored and reliable data on crypto markets

DeFi data and the motivation for Project Atlas

Considering the rapid growth of crypto and DeFi and the associated risks (Graph 1), there is increasing concern about the impact of cryptoassets on the financial system. The size, structural vulnerabilities and integration of DeFi with the conventional financial system could pose risks to financial stability. The implications can differ across jurisdictions. Widespread adoption could undermine monetary policy implementation or limit the effectiveness of capital flow management measures (FSB and IMF (2023)).

Data play a crucial role in measuring economic significance, usage trends and understanding the implications of crypto markets for policymakers. Data gaps are a major concern for regulators. In addition, there is evidence that self-reported data are often unreliable.

Graph 1: Crypto and DeFi grew significantly, but are prone to risk and data manipulation

A. Market size of crypto and DeFi[1]

B. Terra and Luna coins dropped in value rapidly

C. Estimated actual volume is much less than reported[2]

[1] End-of-week values. Categories comprise the seven largest stablecoins, 59 DeFi coins and 64 other cryptocurrencies. DeFi coins correspond to cryptocurrencies issued by DeFi platforms and with a market capitalisation-to-total value locked ratio smaller than 50, as reported by DeFiLlama. For more details, see BIS (2023). [2] Based on data from 8 August 2023. The reported volume is based on data from CoinGecko, and the estimated actual volume is based on data from Cryptocompare. The estimated actual volume is based on a subset of crypto exchanges deemed more reliable.

Sources: CoinGecko; CryptoCompare; BIS (2022); BIS (2023).

Despite the plethora of data on crypto and DeFi markets, why should central banks employ custom platforms? There are several reasons:

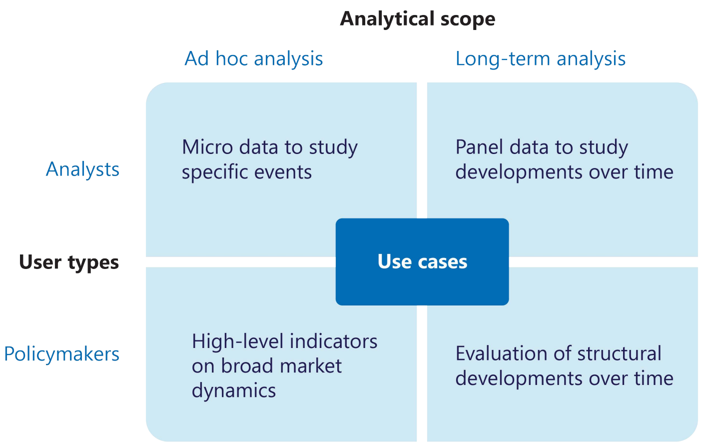
- Crypto data are often **not tailored to central bank needs**. Many data providers focus on information relevant to trading and transaction monitoring. There is a dearth of data tailored to central banks and financial regulators supporting the analysis of the macroeconomic relevance of crypto markets and potential implications for financial stability. Micro data allow central banks to conduct structural analysis, for example relating cross-border flows of crypto to balance of payments data.

- Collecting data on crypto and DeFi **differs from common data collection practices in the traditional financial system**. Collecting public ledger data and voluntarily reported statistics requires a scalable infrastructure to make data easily accessible. Crypto and DeFi data are distributed over a vast network of protocols, market actors and jurisdictions. Processing the data is complex and computationally demanding. Atlas transforms and stores crypto data in a scalable cloud platform, offering a starting point for comprehensive data analysis based on a transparent methodology.
- Analysis of crypto markets needs to account for its unique characteristics. Atlas helps **build up technical and analytical capabilities**, so that central banks can react to new developments in highly dynamic markets. Deeper insights into current data can support developing common methodological standards and conceptual definitions. For example, deriving cross-border flows could be enhanced or complemented by reported data.
- Even though blockchain data are openly available, **methodologies for data aggregation are often a black box**, preventing researchers from engaging in deep-dive analysis and regulators from benchmarking data across sources. By employing granular data and combining different data sources, Atlas can help identify inconsistencies across data sources and why they may arise.
- Data are prone to **manipulation and misreporting** as actors have incentives to exaggerate numbers. For example, crypto exchanges exaggerate trading volumes to move up in rankings. Atlas offers a tool to vet data and calculate statistics based on granular data. Micro data support sensitivity analysis and a clear understanding of data sources. The approach can help inform regulatory initiatives.

Project Atlas creates an infrastructure to collect granular and aggregate data relevant to central banks' mandates, from public and proprietary sources. The project provides a starting point for incorporating further data sources and enables holistic analysis of existing data.

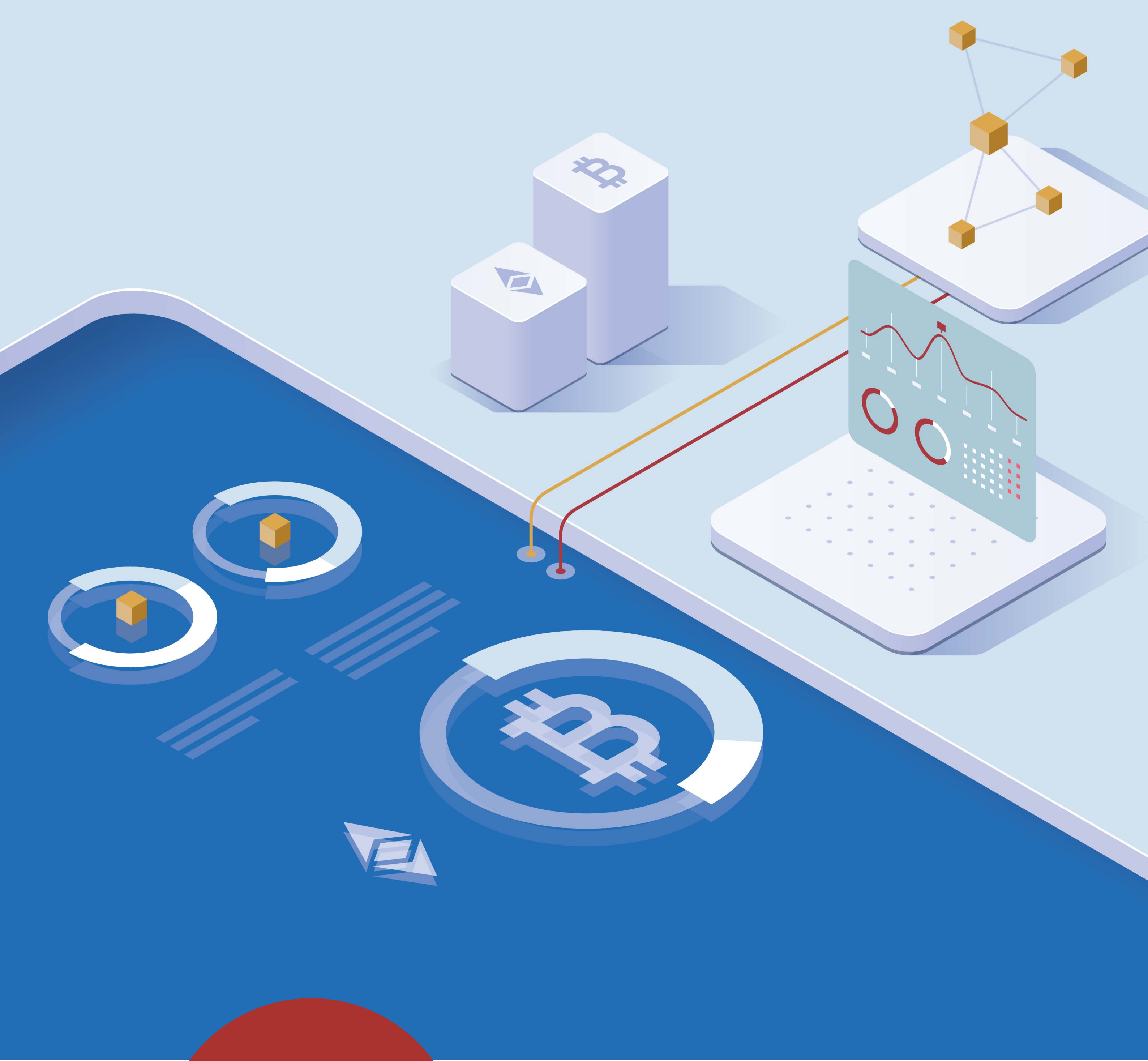
Atlas can cover different use cases, targeting central bank analysts and policymakers. Graph 2 differentiates use cases based on the target audience and the type of analysis. Atlas can make micro data available to researchers to study developments rigorously over time or to analyse specific events ad hoc. For example, researchers could investigate the effect of macroeconomic variables on cross-border flows in crypto or the effect of disruptions in DeFi. At the same time, the platform can serve policymakers via indicators useful for ad hoc analysis and evaluations of broad developments over time. This could include the evolution of inter-exchange flows to evaluate their importance over time or for a given jurisdiction.

Graph 2: Use case overview



Project Atlas targets use cases to serve policymakers as well as analysts, economists and researchers. The analytical scope may concern ad hoc analysis of specific events or long-term analysis to evaluate developments structurally.

Source: Project Atlas.



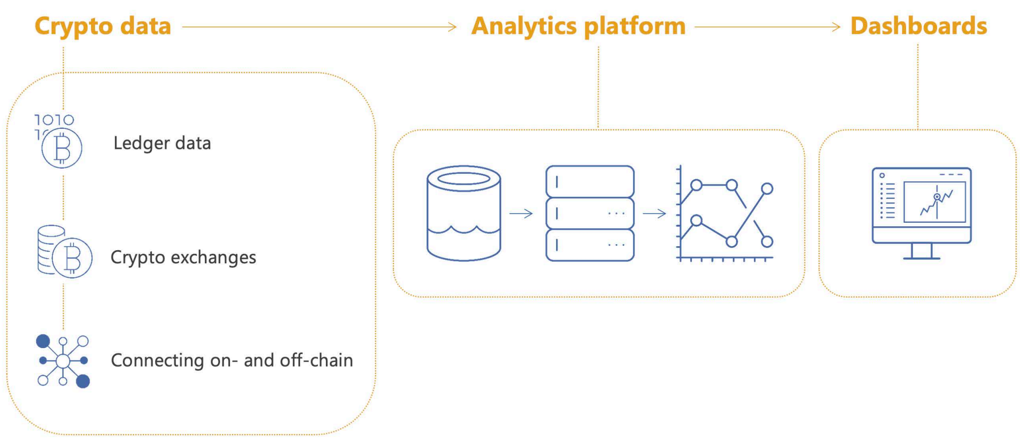
3 Architecture of the data platform

Project overview

Project Atlas consists of data ingestion, the data platform and visualisations in dashboards. The data platform and pipelines lie at the project's heart. The pipelines transport (granular) raw data to feed the analytics platform and generate insights using dashboards (Graph 3). The platform is set up with scalability in mind. It is implemented as a cloud sandbox with the ability to adapt IT resources as needed, allowing users to access the outputs in an analytics environment.

Project Atlas supports two main applications. One is to prepare statistics from raw data for the Atlas dashboards, and the other is hosting structured granular data for users to perform their own analyses. As part of a first PoC, the dashboards focus on capital flows of bitcoin across borders (see Section 5). The data proxy capital flows via on-chain flows between crypto exchanges using attribution data and mapping the location of these exchanges.

Graph 3: Project overview



Source: Project Atlas.

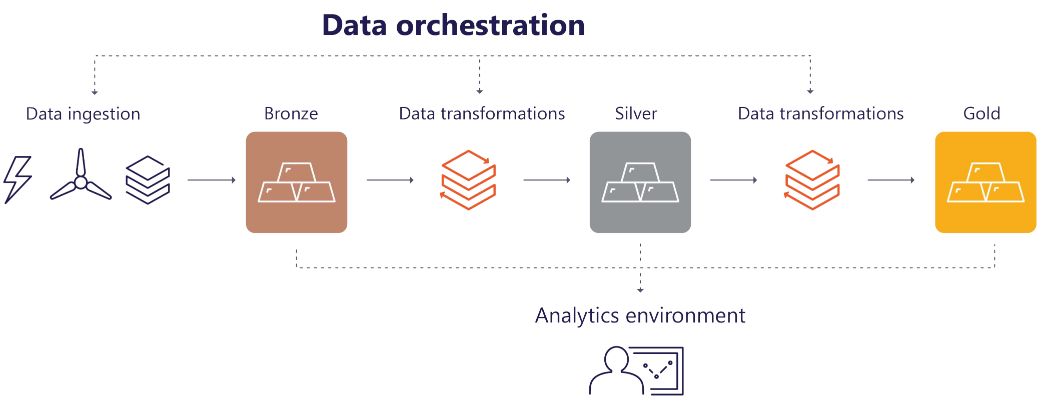
Data ingestion and processing

Given the velocity and variety of protocols and data sources within crypto and DeFi markets, cryptoasset analytics has become a "big data" problem. It requires a corresponding scalable data storage and processing infrastructure. Adding to the complexity is the speed at which crypto markets evolve and adopt new standards. Computing network abstractions such as those representing capital flows and calculating aggregate statistics based on granular data is a computationally resource-intensive task.

The Atlas platform builds and maintains data pipelines to ingest and transform on- and off-chain data into a data lake hosted in a cloud sandbox (Graph 4). The platform is configured and chosen to support the loading of data from any source in their native data formats: large amounts of on-chain data and granular or aggregated off-chain data. The processing logic of the platform integrates diverse data sources, which are enhanced and fused together, enabling adaptable data analysis.

Data are structured in layers in a medallion architecture. The bronze layer ingests the raw data and is equivalent to a landing zone for incoming data. The silver layer includes transformed data that are formatted, cleaned and normalised. The gold layer includes granular and aggregated data in line with the data model specifications that are ready to use for analysts and researchers and serve the dashboards. The logic pushing data to the gold layer includes renaming and reformatting variables and data tables.

Graph 4: Architecture of data pipelines and processing



Data are structured in layers in a medallion architecture. Depending on users' needs, the analytics environment can access data from the bronze, silver or gold layer.

Source: Project Atlas.

The cloud infrastructure is implemented on the Microsoft Azure platform and utilises components that are either widely available or are easily replaced. ETL (extract, transform, load) pipelines transform and process the incoming raw data into a common data structure and load them into the data lake.[4]

[4] Data files are stored in Parquet format and manipulated using Apache Spark. As Spark is open source, the ETL logic has the potential to be applied to any infrastructure that utilises Spark. On top of this structure, the platform employs Hive schemas which allow users to query data using familiar concepts such as SQL syntax. A metadata-driven approach facilitates data management.

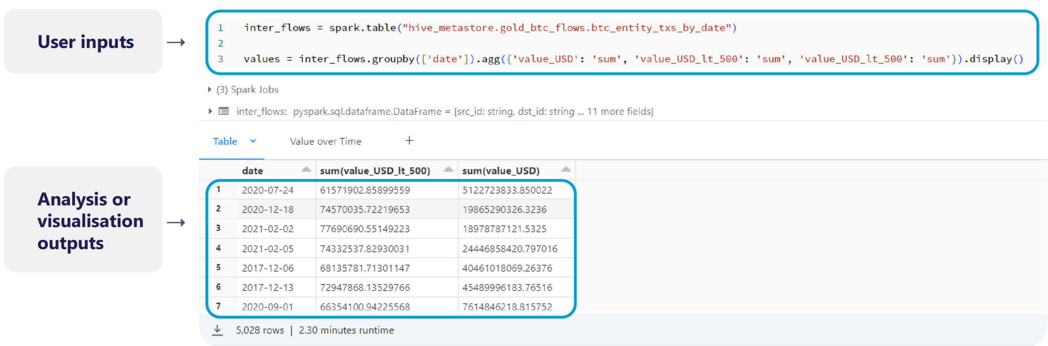
Analytics environment

The analytics environment offers familiar tools and a scalable platform for advanced users. Users could be, for example, economists, researchers or market analysts. The platform is built with Databricks at the centre of its analytics environment, which provides a collaborative environment for processing and analysing big data using Spark. Apache Spark, an open source distributed computing framework, is the platform's engine, offering various features and tools to facilitate data engineering, data science and machine learning workflows.

Databricks provides a notebook interface for interactive coding, targeting researchers and analysts (Graph 5). Data engineers can use the option to automate pipelines in workflows. It integrates with popular programming languages and libraries, allowing users to leverage their existing skills and code. Users work in notebooks to run code in syntax many data analysts are familiar with, like R and SQL.

Users can access the data stored in different layers. Data descriptions include the source of data, the context of the data (such as bitcoin flows or entity master data), the variable names and formats, and sample data. Users can flexibly merge data sources based on their requirements and utilise advanced analytics capabilities within the platform. Additionally, the platform facilitates the creation of visualisations and the export of aggregated data to local environments.

Graph 5: Example of a workspace notebook in the analytics environment



Source: Project Atlas.

Dashboards

Building on the analytics environment, Atlas also provides an instant overview of indicators which are meaningful for policymakers. Pre-calculated indicators serve a front-end web application that showcases intuitive and interactive dashboards, providing users with a comprehensive view of key metrics and trends as well as drill-downs for data exploration. For example, users can investigate cross-border flows for specific jurisdictions in a globe visualisation. The content of the dashboards is described in Section 5.

The web application provides an intuitive and informative interface for users to explore and analyse data efficiently. The web application has been implemented in-house, which enables visualisations to be tailored to different use cases and data types.[5] The dashboards allow users to interact with the data, apply filters and drill down into specific details.

Data journey on the platform: processing on-chain data

The example of Bitcoin transaction data illustrates the data stream from raw data to insights. To gather on-chain data from the Bitcoin network, Atlas operates a dedicated blockchain node. This process entails implementing additional security measures and restricting inbound access to ensure the platform's security. Software such as the Bitcoin client relies on open source software that is updated frequently. Running data ingestion pipelines requires updating clients and adapting the corresponding data pipeline in case of need. Open source software is employed for data ingestion.

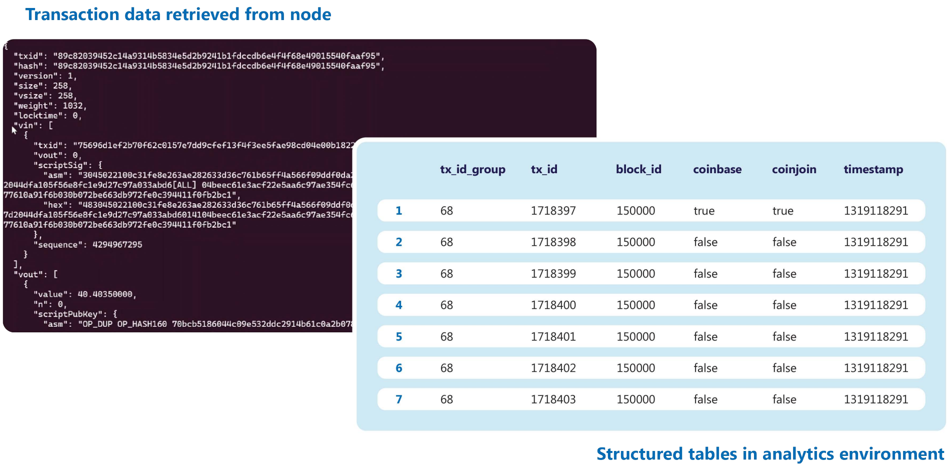
Bitcoin data from the node is extracted in JSON format, which needs further processing to conduct data analysis (Graph 6). Arriving in the bronze layer (Graph 3), the Bitcoin data are transformed into structured tables, including information such as the inputs and outputs of a transaction. For example, these tables contain the block in which the transaction was included and whether or not the transaction was a so-called coinbase transaction, which rewards miners. For the silver layer, Atlas further transforms the data, where possible, into a format familiar to many analysts and akin to payment system data, with a sender, receiver, transaction value and time stamp, alongside more detailed information about the transaction. For the gold layer, the processing engine aggregates the data in different ways, eg the total transactions at a daily level.

Off-chain data complement on-chain data directly or indirectly. Attribution data linking addresses to entities in the network supplement the on-chain data, adding layers that feed into the aggregation of cross-border flows. Other data, such as trading volumes, serve for data comparison.

Extracting blockchain node data and processing them enhances analytical capabilities and allows expertise in technical domains to be built up, as Atlas relies heavily on in-house development. This approach will allow for various use cases, including real-time analysis, although it may necessitate adjustments and additional operational resources.

[5] Implementing these dashboards involves leveraging modern single-page application web frameworks – in this case, React – and versatile data visualisation libraries like D3.js. The application integrates with data sources, fetching and processing the data to generate near-time visualisations.

Graph 6: Transforming raw data into structured table formats and aggregations



	tx_id_group	tx_id	block_id	coinbase	coinjoin	timestamp
1	68	1718397	150000	true	true	1319118291
2	68	1718398	150000	false	false	1319118291
3	68	1718399	150000	false	false	1319118291
4	68	1718400	150000	false	false	1319118291
5	68	1718401	150000	false	false	1319118291
6	68	1718402	150000	false	false	1319118291
7	68	1718403	150000	false	false	1319118291

View of raw data extracted from the Bitcoin node and transformed structured data.

Source: Project Atlas.

4 Data sources and first proof of concept

Data sources

Given the intricate structure of crypto and DeFi markets (see annex for an overview), a comprehensive analytics platform must fuse data gathered from multiple sources. Atlas uses a big data approach to incorporate high-volume and high-velocity data in various formats. Data come from two sources: on-chain data, which stem from public ledgers, and off-chain data, which are reported by market actors or more generally any data other than that extracted from ledgers.

Historical transactions occurring on public, permissionless blockchains are recorded on the public ledger and are visible to everybody. These data are traceable and tamper-proof. On-chain data can provide insights into the distribution of coins, how long addresses hold coins, transaction sizes and throughput in the network. In addition, on-chain data can show the network topology. Atlas extracts on-chain data directly from blockchain nodes. Additionally, the platform imports ledger data and aggregate flows provided by Iknaio Cryptoasset Analytics, which builds on the open-source Graphsense platform.[6]

From Bitcoin-like UTXO (unspent transaction output) model ledgers, one can gather blocks, transactions, addresses and additional meta information.[7] However, many transactions occur off-chain because of greater throughput and cheaper transactions. Such transactions include payments or transactions between users of the same crypto exchange. Off-chain transactions do not occur on the ledger and include centralised intermediaries as transactions settle on the books of market actors. By default, there is no public record for such off-chain transactions. Atlas uses reported trading volumes from crypto exchanges for comparing off-chain and on-chain activity. A mapping file created manually establishes connections between data points, bridging the gap between different data sources. This process entails matching different exchange identifiers and adding country information on the location of crypto exchanges.

The platform is flexible to incorporate other data sets in the future based on user demands. This could include more data on exchanges, data on the user base of crypto exchanges or data related to public sentiment.

[6] See graphsense.info/.

[7] For an overview of the technology and economics for the case of Bitcoin, see eg Böhme et al (2015). Extracting all information from programmable blockchains like Ethereum is more complex. From Ethereum-like account model ledgers, which also provide an execution environment for smart contracts, one can extract information and outputs produced by these programs, such as the source code or the logs and traces of executed program functions and subsequent internal calls between programs. These logs are stored in a separate storage space controlled by the virtual machine environment. The logs and traces are essential for reconstructing token flows (eg stablecoins).

Cryptoasset analytics and entity attribution

Though blockchain data are theoretically transparent, crypto's pseudo-anonymity complicates on-chain data analysis. In contrast to payments data, which include the transacting parties, the real-world identities behind crypto addresses are unknown.[8] This limits analysts' ability to gain practical insights. Akin to payments data, inference is needed to derive the underlying economic rationale.

UTXO ledgers document the complete transaction history in which the inputs and outputs link transactions with each other. However, raw transaction data are typically unsuited for analytical tasks as they do not represent asset flows between entities. Many cryptoasset analyses focus on asset flows between addresses and associated entities and therefore require a network model abstraction, which can be derived from the underlying transaction data set.

Gathering cryptoasset capital flows between entities such as crypto exchanges – ultimately deriving on-chain cross-border flows – relies on several computational steps. The first step is grouping addresses that are controlled by the same actor using clustering heuristics. The second step builds on this clustering, representing the flows between actors that control multiple addresses in a network abstraction. The third step is linking the addresses to the entities they belong to via attribution data to derive the relevant inter-entity capital flows. To assess the economic significance, cryptoasset flows are converted into equivalent values in fiat currencies such as US dollars using historical pricing data. The process for generating capital flows is repeated continuously to integrate the newest data.

For Bitcoin-like UTXO-model ledgers, clustering heuristics identify addresses controlled by the same entity (eg Meiklejohn et al (2016) and Tasca et al (2016)).[9] A user can create and control an arbitrary number of addresses. Linking and clustering these addresses into a single set, representing the real-world entity that probably controls these addresses, is an essential task in cryptoasset analytics. A standard approach is so-called co-spent heuristics, also called multiple-input heuristics. It assumes that inputs jointly spent in the same transaction are controlled by the same entity that must possess the corresponding private key for signing these inputs. While this method has proven very effective in practice, a known, possible source for false positives are so-called CoinJoins, which can be identified and filtered before applying those heuristics (eg Stütz et al (2023)). Other clustering heuristics, known as change heuristics, rely on identifying change addresses in the transaction outputs. Since change address properties are inconsistent, change heuristics can yield false positives (Möser and Narayanan (2017)), ie the approach in some instances clusters addresses

[8] In some cases, payments data include the sending and receiving entities but not the account holders, for example in the context of payment system data. Importantly, payments data mostly does not include the underlying business reason for transactions.

[9] See eg Meiklejohn et al (2016) and Tasca et al (2016).

not controlled by the same entity. Atlas follows a conservative strategy by relying only on multiple-input heuristics and identifying CoinJoins.

With clustering heuristics, it is possible to construct the entity network, representing asset flows between address clusters probably controlled by the same real-world entity. A single entity can control several address clusters. Building on the entity network abstraction, blockchain addresses are de-anonymised and linked to real-world entities using public and proprietary information, referred to as attribution data. Attribution data include information on the acting entity, such as the name of a crypto exchange. The strength of the approach lies in combining address clusters with attribution data. One data point that attributes a single address to a real-world entity can identify a large address cluster. This way, the approach can at times even de-anonymise a couple of hundred thousand addresses with a single data point.

However, generating attribution data is usually expensive since it relies on sample interactions with a particular crypto exchange, crawling for published addresses or other more elaborate data-gathering procedures. There are public sources for attribution data (eg walletexplorer.com or etherscan.io). In addition, private companies increasingly offer attribution data as part of their business model. Therefore, comprehensive attribution data are often proprietary information. While the focus often lies on forensics or transaction screening, Atlas employs attribution data focusing on macroeconomic relevance. Atlas employs public attribution data combined with proprietary data provided by Iknaio research, which is also based on third parties that specialise in cryptoasset data or indirectly collected attribution data.[10] The platform updates data from different repositories and can incorporate further sources of attribution data.

Proof of concept: modelling cross-border flows

Project Atlas data use transactions between crypto exchanges in the Bitcoin network, along with the location of those exchanges, as a proxy for cross-border capital flows (Graph 7). Since the jurisdiction is not always discernible for crypto exchanges due to multiple locations or the absence of a registered central location, the estimates of cross-border flows should be considered lower-bound estimates. Given the nature of crypto markets and the lack of global and comprehensive regulatory data collection, the approach represents an effort to estimate cross-border flows under uncertainty.

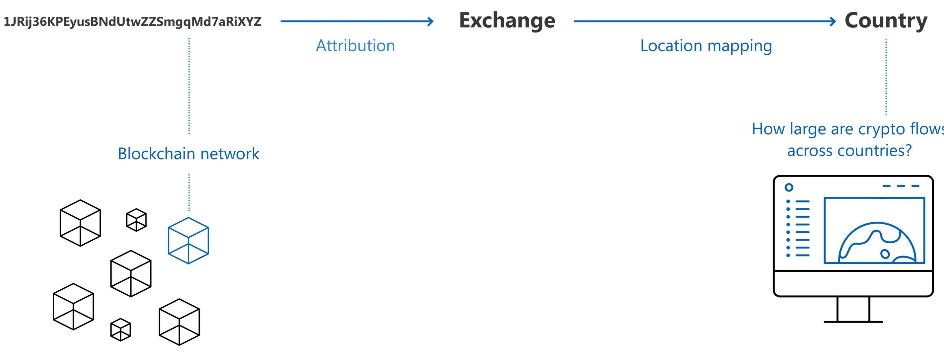
Based on an identified primary location where an entity is registered, the exchange flows are mapped to that country to derive crypto cross-border capital flows (Graph 7). This approach comes with the drawback that flows are based on exchange locations rather than transacting users' locations. However, the approach provides

[10] The aggregation of proprietary attribution data on crypto exchanges is performed by Iknaio Cryptoasset Analytics, while the other parties do not have access to this granular data. The aggregated flows are then included in the Atlas platform (added 16 October 2023).

information on where the entities controlling the addresses are located and what changes occur in their coin holdings.

In future iterations, other country attributions will be explored. This process could include mapping exchanges to countries based on the distribution of the user base across locations.[11]

Graph 7: Deriving cross-border flows based on crypto exchange locations



Source: Project Atlas.

[11] This will still only constitute an approximation of overall user-to-user flows.

5

The Project Atlas dashboards

The Project Atlas dashboards

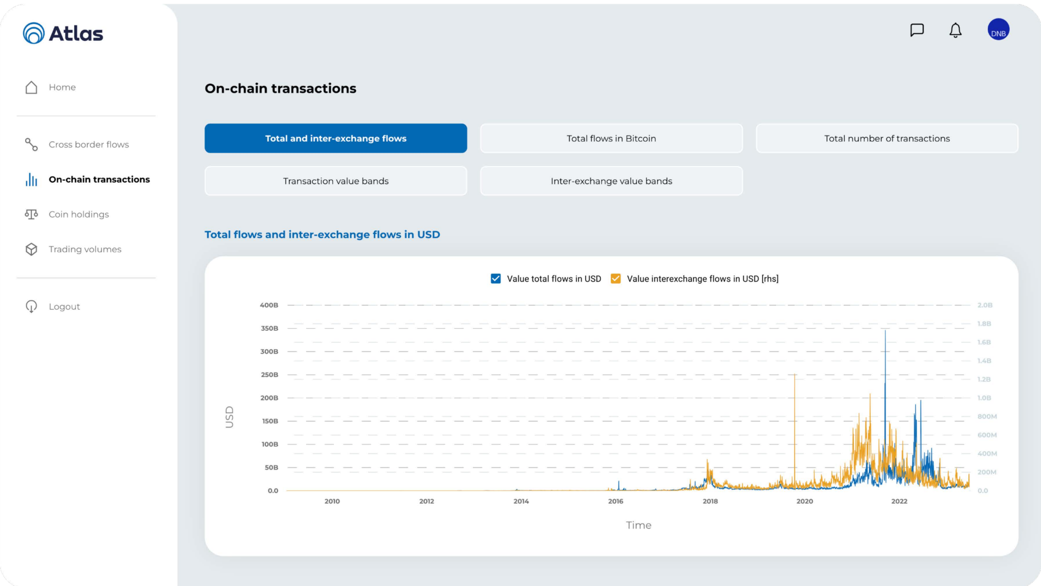
The Project Atlas front end showcases dashboards which visualise the results of data aggregation and analysis that interest policymakers.

On-chain transactions

The left-hand menu includes different categories of data visualisations with panels of graphs under each (Graph 8). Within the visualisations, users can drill down further into the data. The categories will grow with the expansion of data sources beyond the first PoC.

The on-chain transactions category shows on-chain flows in different breakdowns as time series. The dashboards show total flows in the unit of cryptoassets converted to US dollars. In addition, the number of transactions in the network can provide insights into the adoption and relative importance of crypto markets. Graph 8 depicts overall on-chain volumes at a daily frequency. As a subset of total transactions, the identified flows between crypto exchanges are presented as aggregate numbers over time. Relative to total flows, inter-exchange flows are relatively small. However, they are substantial in economic terms and growing over time. The inter-exchange flows are a lower-bound estimate relying on the methodology described earlier. They provide insights into how relevant new market actors in crypto ecosystems are.

Graph 8: Dashboard menu and on-chain transaction visualisation



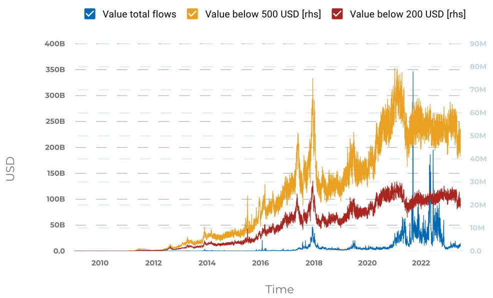
Visualisation showing total on-chain transfers in the Bitcoin network and the subset of transactions occurring between identified crypto exchanges.

Source: Project Atlas.

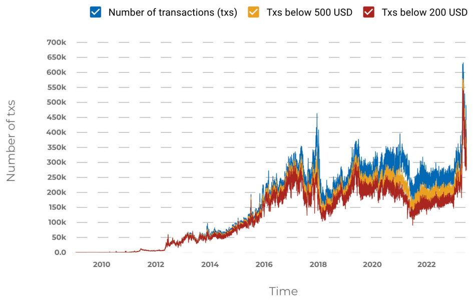
Drilling down into total transactions, the value bands into which transactions fall can indicate how relevant crypto transactions on-chain are for retail payments and remittances. Such payments typically fall into smaller value bands. However, further criteria would be needed to reliably identify the purpose of such transfers. Graph 9 distinguishes transactions in Bitcoin equivalent to a value below $500 and below $200. The value of these transactions is almost negligible compared with total volumes, showing that on-chain transactions are highly skewed towards large-value transfers. However, they make up most transfers when looking at the number of transactions.

Graph 9: Transaction value bands

A. Transaction value bands – value



B. Transaction value bands – number



Graphs showing transaction value bands in terms of the value and number of transactions, respectively. Transactions below $500 and $200 in panel A on right-hand scale.

Source: Project Atlas.

Exchange-based cross-border flows

There has been much interest in cross-border crypto flows.[12] Visualising the data can give a global overview of flows, how they evolved, and their magnitude and concentration. Policymakers and supervisors can zoom in on flows and gain new insights relevant to their jurisdiction.

Based on the inter-exchange flows, Atlas derives cross-border flows based on the location of crypto exchanges where that information is available. Cross-border flows based on attribution data and exchange locations are a vast data set. It includes daily bilateral inflows and outflows for individual countries with registered exchanges. However, on-chain flows form only one part of cross-border crypto transfers and are a lower-bound estimate. Actual flows are likely to be higher, but given the complexity and nature of crypto markets, comprehensive data are non-existent.

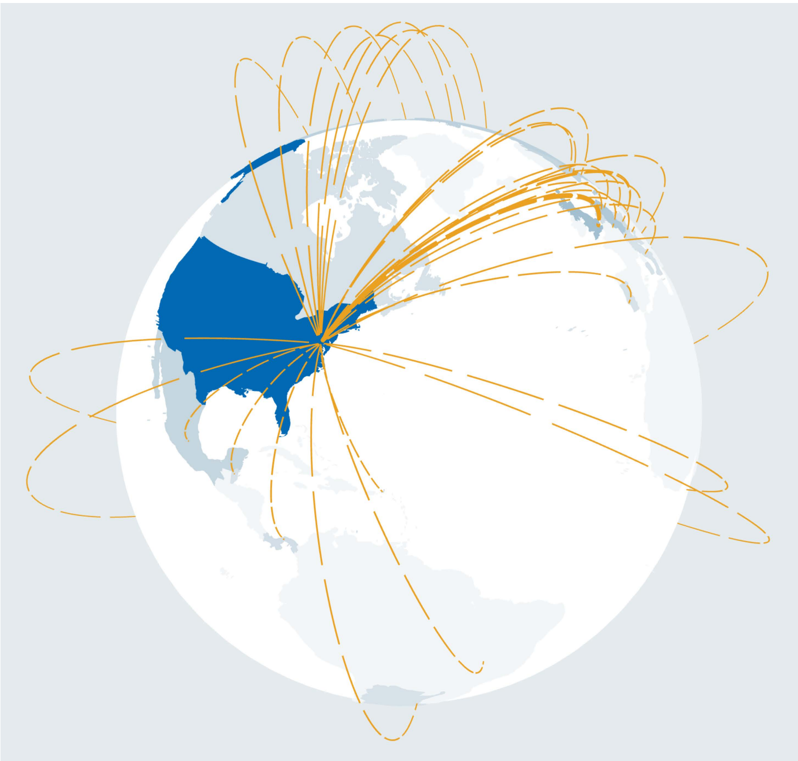
An initial analysis of data collected by the platform indicates that cross-border flows are substantial in economic terms and unevenly distributed across geographical regions.

[12] For example, Graf von Luckner et al (2021) use data from centralised peer-to-peer Bitcoin markets to investigate cross-border flows.

For visualising and allowing drill-down possibilities in a user-friendly way, Atlas shows cross-border flows on a globe at a yearly frequency (Graph 10). Users can rotate the globe and click on individual countries for which flows of crypto exchanges are available. Based on country selection, animated arrows show inflows and outflows and accentuate flows according to their relative importance. The absolute amounts in US dollars are visible in a sub-menu. The visualisation enables policymakers and analysts to gain insights into individual jurisdictions through time at a glance. The globe implementation is flexible, allows new features to be incorporated and may also serve to visualise other cross-border flows on crypto, DeFi and beyond.

Cross-border flow size and distribution are particularly relevant for central banks in the context of cross-border payments, economic analysis and balance of payments statistics. Transfers of cryptoassets may represent a significant part of cross-border transfers for some jurisdictions. At the same time, crypto flows are typically not fully captured in balance of payments and other data. While relatively small compared with international capital flows in traditional financial markets, crypto flows can have different implications across jurisdictions.

Graph 10: Globe visualisation of cross-border flows



Data are based on inter-exchange flows in the Bitcoin network. The varying thickness of lines reflects the relative significance of values.

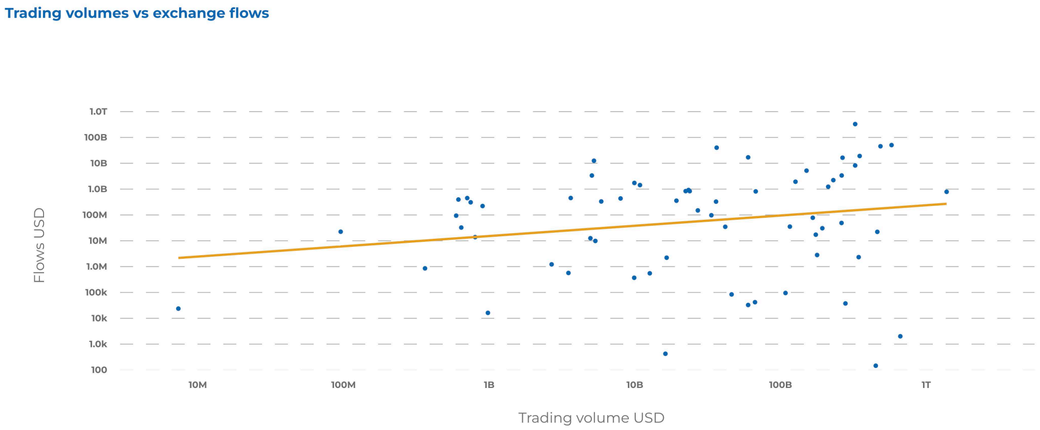
Source: Project Atlas.

On-chain and off-chain comparison

Atlas uses scatter plots to visualise the relation between on-chain and off-chain activity. While there is no causal relationship, higher off-chain activity should correspond to higher on-chain activity. For example, with a larger number of users trading on a crypto exchange, the exchange should exhibit higher coin holdings and increased on-chain flows.

The scatter plots compare the two dimensions at an exchange level. Graph 11 plots the on-chain inflows and trading volumes on a yearly basis. Given the highly skewed values, the data are shown on a log scale. One data point in the plot represents one exchange with its associated on-chain flows and reported trading volumes in a given year. Given the uncertainty in identifying on-chain activity and the fact there is no strict correlation, a mismatch between on- and off-chain activity should not be regarded as proof of misreporting. However, mismatches with high reported trading volumes but low on-chain activity can indicate cases of inflated trading volumes in some instances. Such cases could call for further analysis. To further compare reported trading volumes and on-chain activity, Atlas uses coin holdings of exchanges.

Graph 11: Scatter plot, trading volumes and on-chain flows



Data for 2022. One data point represents trading volumes of a crypto exchange on the horizontal axis (log scale) and the value of on-chain inflows for that exchange on the vertical axis (log scale) in a given year. Data exclude exchanges for which no on-chain flows or no trading volumes were identified.

Sources: Cryptocompare; Project Atlas.

6 Conclusion and next steps

Conclusion

Central banks need to gain first-hand knowledge of crypto and DeFi and the risks and opportunities they present to the financial system. Currently, crypto and DeFi markets are challenging to map because they are opaque and market participants often do not comply with existing regulations or operate outside the regulatory remit. Reliable and granular data are difficult to obtain. Project Atlas provides a starting point for solving some of these challenges.

Atlas serves researchers and analysts as well as policymakers, supporting structural analysis and offering a user-friendly tool which yields intuitive insights. It allows central banks to assess the significance of crypto and DeFi markets and can enable policymakers to make better informed decisions.

Analytics is only as powerful as the amount and transparency of data available. Inspecting, identifying, understanding and visualising data allows users to gain valuable insights that would otherwise be hidden. While there are abundant public data on crypto and DeFi markets, extracting insights from them requires solid understanding of them. Cryptoasset analytics can empower better predictive modelling for assessing systemic risks, thereby enhancing the safety and soundness of the financial system.

The output of Project Atlas could serve as a starting point for preliminary assessments and inform the drafting of data reporting requirements and regulation of crypto market actors. Based on a clear and transparent methodology, Atlas enhances central banks' technical capabilities and strengthens the understanding of crypto and DeFi data.

Next steps

Atlas offers a starting point for building a comprehensive cryptoasset analytics environment for regulators and the central banking community that is expandable to new data sources and analytical methods. Several centralised crypto exchanges today make information on the cryptoasset addresses they control public – under the label "proof of reserves" – for transparency purposes. In the future, such disclosures could be used for novel approaches to data-driven supervision.

Incorporating further data sources will be a crucial priority in the next phase. Extracting and analysing data from a blockchain node in the Ethereum network is envisaged. As DeFi markets expand, this is a natural step for further analysis. Furthermore, the project team will seek feedback from the central banking community on which statistics help them fulfil their mandates.

The data platform and dashboards will be made available to test users in order to gather feedback on functionality and data needs. This will inform the development of additional features and visualisations.

References

Aramonte, S, W Huang and A Schrimpf (2021): "DeFi risks and the decentralisation illusion", *BIS Quarterly Review*, December, pp 21–36.

Auer, R, M Farag, U Lewrick, L Orazem and M Zoss (2022): "Banking in the shadow of Bitcoin? The institutional adoption of cryptocurrencies", *BIS Working Papers*, no 1013.

Auer, R, B Haslhofer, S Kitzler, P Saggese and F Victor (2023): "The technology of decentralized finance (DeFi)", *CEPR Press Discussion Paper*, no 18038.

Bank for International Settlements (BIS) (2022): "The future monetary system", *Annual Economic Report 2022*, June, Chapter III.

——— (2023): *The crypto ecosystem: key elements and risks*, report submitted to the G20 Finance Ministers and Central Bank Governors, July.

Böhme, R, N Christin, B Edelman and T Moore (2015): "Bitcoin: economics, technology, and governance", *Journal of Economic Perspectives*, vol 29, no 2, pp 213–38.

Cong, L, X Li, K Tang and Y Yang (2023): "Crypto wash trading", *Management Science*, forthcoming.

Deutsche Bundesbank (2021): "Crypto tokens and decentralised financial Applications", Monthly Report July 2021.

Financial Stability Board (FSB) (2022), *Assessment of risks to financial stability from crypto-assets*, technical report to the G20.

Financial Stability Board and the International Monetary Fund (FSB and IMF) (2023): *IMF-FSB synthesis paper: policies for crypto-assets*, September.

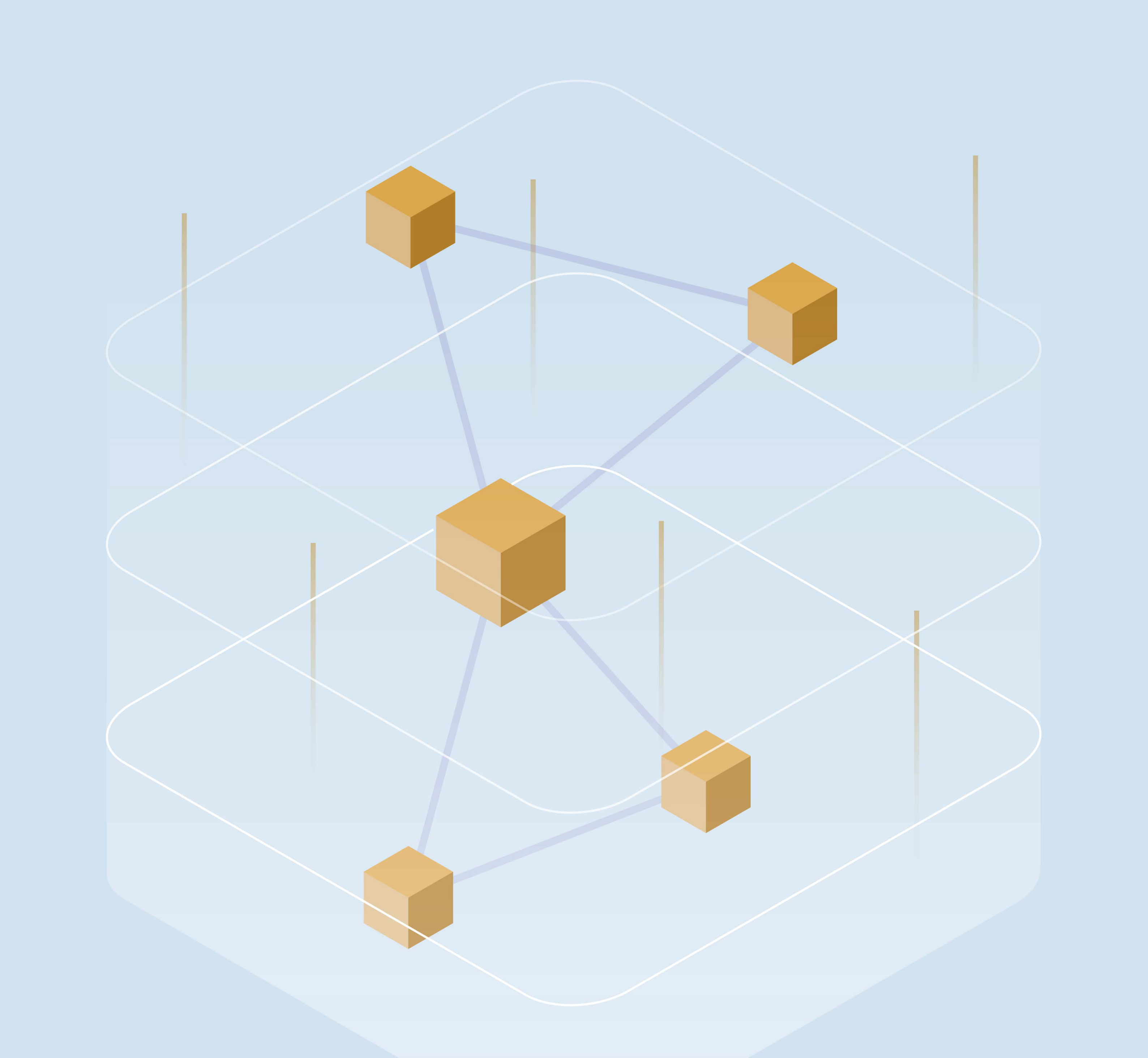
Graf von Luckner, C, C Reinhart and K Rogoff (2021): "Decrypting new age international capital flows, *NBER Working Paper*, no 29337.

Meiklejohn, S, M Pomarole, G Jordan, K Levchenko, D McCoy, G Voelker and S Savage (2016): "A fistful of Bitcoins: characterizing payments among men with no names", *Communications of the ACM*, vol 59, no 4, pp 86–93.

Möser, M and A Narayanan (2022): "Resurrecting address clustering in Bitcoin", in I Eyal and J Garay (eds), *Financial cryptography and data security*, *Lecture Notes in Computer Science*, vol 13411.

Stütz, R, J Stockinger,P Moreno-Sanchez, B Haslhofer and M Maffei (2023): "Adoption and actual privacy of decentralized CoinJoin implementations in bitcoin", Proceedings of the 4th ACM conference on advances in financial technologies.

Tasca, P, A Hayes and S Liu (2018): "The evolution of the Bitcoin economy: extracting and analyzing the network of payment relationships", *The Journal of Risk Finance*, vol 19, no 2, pp 94–126.



Annex

Crypto markets and the building blocks of DeFi

Cryptoassets and DeFi applications have evolved as a novel ecosystem. As they are global by nature, transactions cross national borders. Initially used only by a small, tech-savvy group, cryptoassets have begun to reach a broader audience. As a result of the growing ecosystem, new actors emerging in this new, largely unregulated ecosystem become a source of concern for regulatory institutions and supervisory bodies. This is especially true in the light of recent events such as the bankruptcy of the crypto exchange FTX and the collapse of the Terra (Luna) stablecoin. At the same time, some aspects of these new markets could inspire new applications in traditional financial markets.[13] Not least, the possibilities of new technologies and the potential for enabling efficient global payments inspired many central banks to investigate central bank digital currencies (CBDCs).

The emergence of Bitcoin in 2009 marked the birth of crypto, a system that allows users to transfer cryptoassets on a permissionless and decentralised public ledger (the blockchain), relying on distributed ledger technology (DLT). The blockchain is called permissionless because anyone can become a validator, ie participate in validating transactions. The record-keeping is performed by multiple validators using a consensus mechanism, rather than trusted intermediaries such as banks, as is the case in the traditional financial system Therefore, the system is considered to be decentralised. The history of all transactions is recorded by the blockchain. While transactions are tied to specific wallets, the true identities of the owners of the wallets (ie the parties behind transactions) remain hidden behind cryptographic digital signatures.

While Bitcoin introduced the idea of the blockchain to the public, many other blockchain implementations have since built on its design. The most notable is arguably Ethereum, with its associated coin ether. Ethereum and other blockchain implementations allow for programmability and hence the use of smart contracts. Smart contracts are self-executing bits of code that trigger an action if some pre-specified conditions are met. They run on a network of computers to automatically manage financial transactions. Implemented on top of DLT, DeFi does not require banks or other traditional centralised intermediaries. The DeFi ecosystem consists of financial protocols implemented as “smart contracts”. By combining multiple smart contracts, users can perform complex transactions. The term “financial lego” illustrates that multiple smart contracts can be assembled or composed so as to offer novel financial services (Auer et al (2023)).

Programmable blockchains underpin the rise of DeFi. For example, users can lend their cryptoassets on lending platforms to borrowers who post other cryptoassets as collateral, collecting interest (Aramonte et al (2022)). So-called stablecoins play a vital role in DeFi. Stablecoins are either asset-backed or algorithmic. Asset-backed stablecoins are pegged to assets or currencies such as the US dollar. Algorithmic stablecoins are supposed to maintain their value relative to the underlying numeraire

[13] One example is the BIS Innovation Hub’s Project Mariana.

via algorithms that automatically adjust supply. The primary use case of stablecoins is to overcome the enormous price volatility and low liquidity of most unbacked cryptoassets. However, the collapse of algorithmic stablecoins due to downward spirals has illustrated the susceptibility to risks.

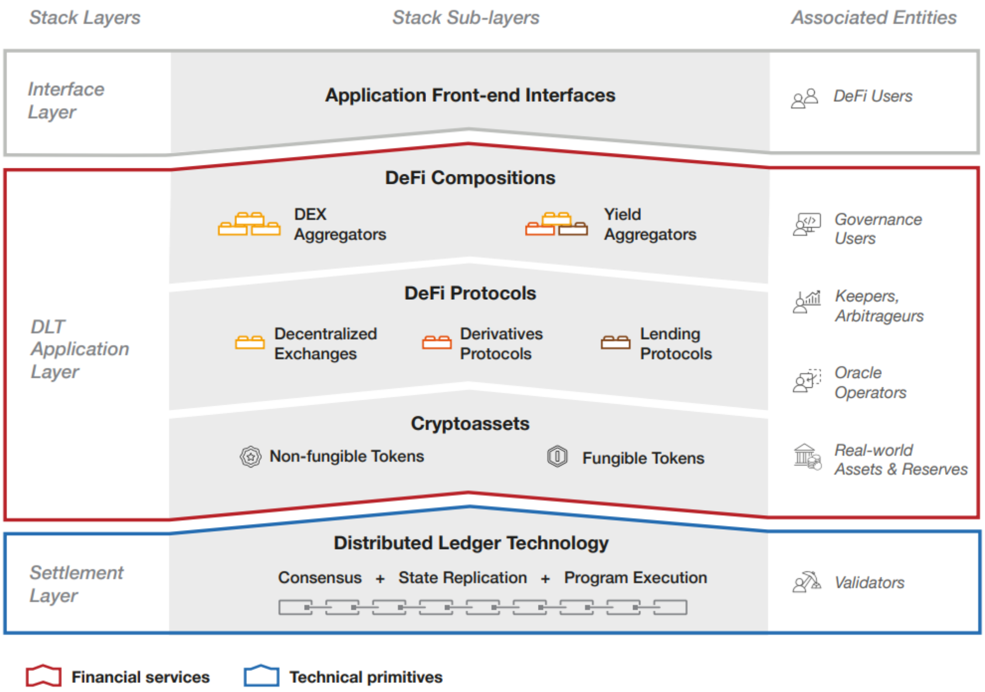
The interface between conventional and decentralised finance can be called centralised finance (CeFi) (see eg Aramonte et al (2022)). CeFi includes crypto-financial companies exploiting off-chain forms of accountability – companies that de facto act as intermediaries between customers and record financial trades in private databases. Two types of CeFi actors play a crucial role in bridging the two ecosystems. First, centralised crypto exchanges provide cryptoasset investors with interfaces to conventional payment systems and provide the fiat legs of crypto-fiat trades, which can act as de facto banks (Böhme et al (2015)). Second, fiat-backed stablecoin providers maintain the peg to the numeraire through reserves.

Atlas explores broad dynamics and the most relevant actors in the ecosystem. Crypto exchanges are a natural starting point, as market activity is highly concentrated among these actors (FSB (2022)). At the same time, research shows that reported data of crypto exchanges, especially unregulated, ones, is often unreliable. Cong et al (2023) estimate that up to 70% of reported trading volumes are wash trades.

The DeFi ecosystem is a composite and fragmented collection of individual subsystems, each built on top of a DLT with specific technical features. Breaking down the broader DeFi ecosystem, Auer et al (2023) develop a DeFi stack reference (DSR) model. The model comprises a settlement, application and interface layer (Graph 12). In the first phase, Atlas focuses on one specific aspect of crypto and DeFi markets: crypto exchanges in the context of Bitcoin on-chain flows and trading. This aspect involves the settlement and DLT application layer and centralised actors providing services based on those layers.

In contrast to the conventional financial system, data-gathering in crypto and DeFi poses new challenges. One can hardly determine in which jurisdictions a service provider operates since services are provided in online pseudo-anonymous networks. Furthermore, there is a lack of provisions on responsibilities and liability. Whereas traditional financial intermediaries are subject to regulation and reporting standards, crypto service providers often operate outside the regulatory remit, hindering the application and enforcement of standards. At the same time, the fact that transaction data are stored in a public ledger enables new types of data elicitation; albeit with higher uncertainty.

Graph 12: DeFi stack reference model



Source: Auer et al (2023).

Contributors

Contributors

BIS Innovation Hub

Calin Simon, Adviser

Jan Paulick, Adviser and Project Lead

Nina Gonschorreck, Adviser

Sally Dubach, Data analyst

Timothy Aerts, Adviser

De Nederlandsche Bank

Henk Esselink, Senior Manager, Steering Committee

Deutsche Bundesbank

Julia Biesen, Senior Manager, Steering Committee

Martin Diehl, Senior Manager, Steering Committee

Bank of France

Alexandre Le Douaron, Senior Manager (Observer)

Marc Fasquelle, Senior Manager (Observer)

European Central Bank

Sjoerd Van der Vaart, International Innovation Manager (Observer)

Vendor

Bernhard Haslhofer, Co-founder and Research Lead

Michel Fröwis, Development Lead

Rainer Stütz, Co-founder and Data Science Lead

Acknowledgements

The authors are grateful to Bénédicte Nolens, William Zhang, Italo Borssatto, Manuel Fangmann, Johannes Rogowsky and Henry Holden for reviewing the report and their incredibly insightful feedback; to Giulio Cornelli, Ulf Lewrick, Anneke Kosse and Ilaria Mattei for very helpful comments and discussions; to Darko Micic and Bernard van den Boom for invaluable technical contributions; and to the project's private sector partner Iknaio Crypto Analytics for their work.

Bank for International Settlements (BIS)

ISBN 978-92-9259-693-4 (online)