Looking ahead: considering the wider repercussions of regulations

Assessing the impact of Open Banking and PSD2 on systems of record

Contents

1. Background and introduction

Established retail banks are facing a major wave of intentional regulatory disruption in the form of PSD2 and Open Banking. Setting aside the impact on their future business models, the technical ramifications for their existing IT systems are potentially huge.

The typical existing and desired situations are illustrated in Figure 1. The key point to note is that post the Open Banking changes, rather than traffic being mediated via the bank's existing channels, it will be injected directly into the core systems of record (at least from a logical perspective).

Clearly, these legislative changes have a technical impact on the banks, and will place a load on their existing systems of record: that is the focus of this paper. The need to migrate workload or generate new workload in response to the regulations will create a large degree of risk. For example, moving just 5 percent of card transaction volumes to Open Banking would double the load on the Faster Payments scheme – something few banks (nor indeed the scheme) could effortlessly accommodate. Similarly, having the constant load of systems polling for transaction data will create an unprecedented load on unprotected systems of record[1].

2. Regulatory Backdrop

Payment Services Directive 2 (PSD2) is European legislation deliberately intended to disrupt the banking industry. PSD2 requires that all European banks (over 4,000 financial institutions) open access to their back-end services for account information and payments to third parties, through open Application Programming Interfaces (APIs). This regulation is intended to transform the financial services landscape by breaking the banks' monopoly on access to payment mechanisms, enabling participation from other sectors to drive innovation.

In the UK, the Competition & Markets Authority (CMA) is driving a similar agenda, mandating that the nine largest banks in the UK, the so-called "CMA9"[2], open access to all payments accounts. The CMA have appointed the Open Banking Implementation Entity (OBIE)[3] (evolved from the

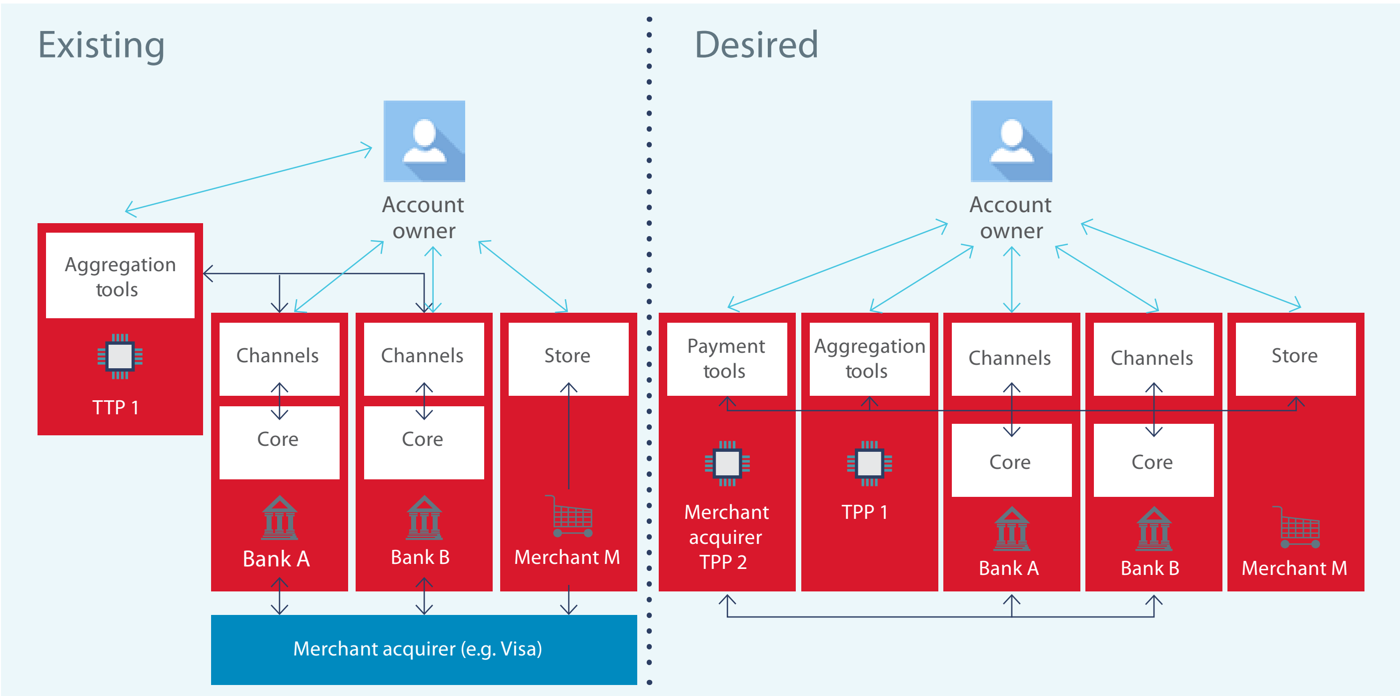

Figure 1: Existing and desired landscapes before and after PSD2 / Open Banking

Open Banking Working Group (OBWG)) to define standards, and run the machinery to implement much of this effort – in effect acting as the enforcement vehicle for much of the PSD2 legislation. Consequently, taking a UK focus, we generally refer to this group of changes under the broad banner of "Open Banking" in this paper.

By the end of March 2017, the initial CMA milestone (CMA1) required that CMA9 deliver Open Data APIs providing product details, ATM and branch locations[4]. The second milestone (CMA2) required the provision of Account Information Services (AIS)[5], and Payment Initiation Services (PIS)[6] by January 2018. The introduction of AIS and PIS enables Third Party Providers (TPPs) to access the customer's account history (including mandates on the account) and to initiate payments on their behalf, with appropriate consent.

The PSD2 and CMA regulations also have an indirect relation to the new General Data Protection Regulation (GDPR). The interrelations between the different regulations include some key differences and tensions including:

- PSD2 and CMA remedies are intended to drive sharing and openness. GDPR is intended to give users control over their data – which greatly complicates the "opening up" of the data.

- PSD2 covers all payment schemes in Europe, and both debit and credit cards. The CMA remedies and associated OBIE standards only cover the UK (so exclude SEPA payments, and non-UK domiciled banks), and don't include card payments.

- PSD2 covers corporate banking, CMA does not.

- OBIE implementation scope for January 2018 is faster payments (Single Immediate Payment) only, with further tranches of scope planned for August 2019 for the delivery of CHAPS and international payment schemes and credit cards.

Despite these differences, European regulators have increasingly looked to the UK (and the OBIE) to lay down the standards that are likely to be adopted in many parts of Europe[7]; placing UK banks, and the CMA9 specifically, at the front of this revolution.

The standards developed by the OBIE are RESTful services[8] – accessed using HTTP(s) as a transport, and with security and non-repudiation techniques wrapped around them. This has become the "new standard" approach for integration and development – replacing approaches such as Service-Orientated Architecture (SOA) and Enterprise Application Integration (EAI). The intent of all this change is to build a new API economy – where services can be composed from freely available APIs to drive new customer value and business models.

3. Workload drivers and projections

3.1. Drivers

It is important to separate new workload generated by Open Banking from workload that is simply originated from a different source (which would be net neutral to the systems of record). The drivers for workload can be plotted on two key axes:

- The **source** of the workload (account information versus payment initiation): Account aggregation scenarios primarily revolve around AIS, and "Pay by Bank" scenarios around PIS. Online banking drives a combination of the two. Each of these triggers a distinct set of downstream interactions with both authorisation services to check their validity, and systems of record to execute them.

- The **nature** of the interaction (directly customer interactive, or background): The Open Banking specs allow interaction both when the customer is directly interactive, and offline where they have provided persistent consent. Within current Open Banking specifications, it is not possible to provide persistent consent for payments, so background load will not come from this space initially. However, data aggregation consent can last up to 90 days, with four interactions a day allowed without the customer present (a limit that is also unlikely to be usefully enforceable). There is nothing within the ecosystem to level out the load of background polling – and hence this is likely to drive significant volume spikes.

Figure 2 shows the different scenarios with a Red, Amber, Green (RAG) categorisation indicating how these are likely to impact back-end systems relative to today.

If we accept that it is reasonable to assume that customer engagement with banking for its own sake is unlikely to be changed by the advent of Open Banking, and that the simplification of existing processes through use of Open Banking is likely to be net neutral in terms of access to systems of record (marked green and amber respectively), then we are left with the following areas of focus:

- Background load driven by **account information polling:** Assuming TPPs wish to ensure that their data is as accurate as possible, and will consequently poll four times a day (the maximum allowed by the standards when the customer is not online), then there are likely to be traffic spikes around 0000hrs, 0600hrs, 1200hrs and 1800hrs when data is brought up to date. While these are unlikely to be instantaneous workload they will be tightly clustered, and likely to come in as quickly as the TPPs are able to process them: there is nothing in the OBIE standards to drive any evening out of load.

- **New customer interactions:** If TPPs deliver applications that aggregate customer accounts across multiple providers, but enable only online banking capabilities, then net new customer transactions will be zero. However, the intent of

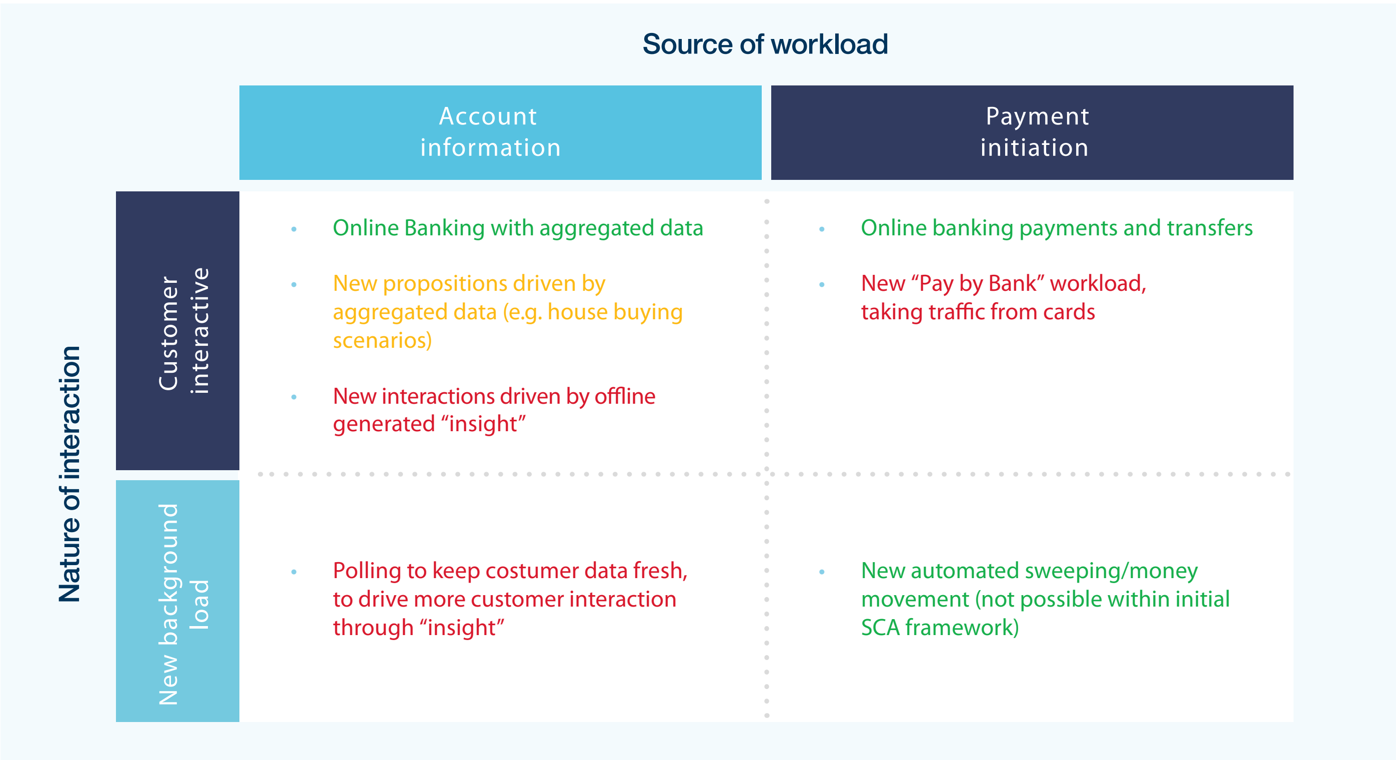

Figure 2: Taxonomy of new volume drivers

the legislation, and the many players in the new market, is to use aggregated data to drive new interactions – and hence drive value for both customers and the company. TPPs will be actively looking to build this category of interaction, to capture customer attention and interaction, and hence drive new workload as hard as they can.

- **New “Pay by Bank” workload:** Pay by Bank – the ability to directly transfer funds from a customer account to another (belonging to either merchants or other consumers in a peer-to-peer transfer scenario) – will be extremely attractive to merchants looking to reduce their card acquirer fees (which can be up to 3 percent per transaction for a credit card payment). The real challenge posed by moving workload from card schemes to Faster Payments is the volatility of the workload. Cards networks are designed to cope with huge spikes in workloads, driven by events such as charitable events and appeals. Such workloads are not present in the existing Faster Payments world, and the first application that provides Pay by Bank as a settlement option for such a scenario will come with unpredictable workload. With cards carrying well over an order of magnitude more transactions per month than Faster Payments this threat could be overwhelming.

3.2. Uptake

Uptake for Open Banking services – both AIS and PIS – is impossible to predict, but is likely to be driven by the first “killer app” to emerge in each of these domains. However, we can identify the key factors that will drive uptake:

- The number of **payment account customers** (i.e. those in immediate scope for PSD2 / Open Banking).

- Percentage of payment account customers who are **active digital customers** (either online or mobile) and hence most likely to engage with Open Banking based propositions, apps and/or services. It is reasonable to assume that, until it reaches ubiquity, these will be the users of Open Banking. An estimated 50 percent of customers fall into this category.

- **Barriers to adoption** for customers. In its earliest incarnations, the consent process for Open Banking is more complex and intrusive than for credit card applications, and potentially more intimidating in terms of presenting terms and conditions around data sharing. This is likely to present a psychological barrier to adoption for many, both from a fear of security issues and from the additional friction introduced to the process. In this paper, we assume a conservative uptake of 5 percent in the initial year, accelerating as adoption and market value grows through the network effect, consumer familiarity with the model and ongoing simplification of the security model, growing to near ubiquity at 80 percent by the end of year five.

- **Current account information profile.** For an active digital customer, this is driven primarily by the number of logons per day, and number and type of actions per logon. We assume each digitally active customer logs in every other day, performing a balance check and a query of the latest transactions on their account with their online banking app, equivalent to an average of one query per day.

- **Number of services used.** Other than the initial and ongoing consent processes there is nothing to prevent a customer signing up to multiple services. We assume that each customer that engages with Open Banking will sign up for, on average, two services. This is likely to increase in time alongside diversity of services, but we will assume it is stable.

- **Move from cards to payments**. Possibly the hardest element to predict, for the purposes of this study we assume a 5 percent shift away from card to Pay by Bank. This 5 percent represents less than the current annual growth rate of card schemes (which stands at around 6.8 percent, with the online element of cards being up 26 percent – and a key area where Pay by Bank is likely to be targeted).

3.3. Volumetric Projections

While they may target several common systems of record with their aggregate workload (e.g. the accounting platform), the AIS and PIS scenarios are driven by very different dynamics, so we examine their projections separately.

3.3.1. Account Information

Taking our assumptions above, at the end of five years, 80 percent of the digitally active customer base will be using AIS (Figure 3).

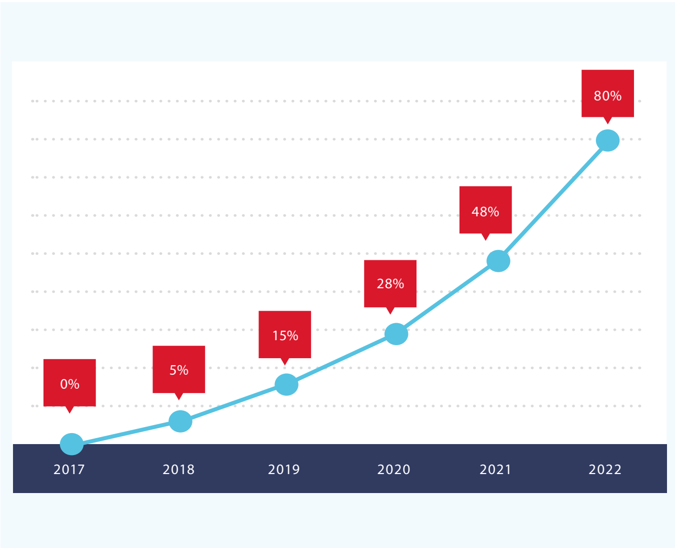

Figure 3: Projected Open Banking AIS uptake

Two active Open Banking services, polling four times a day, making a conservative two calls (for balance and recent transactions) each time, drives 16 net new queries to systems of record for each Open Banking adopter within the bank, compared to one (two queries, every other day) for an existing digital user (Figure 4).

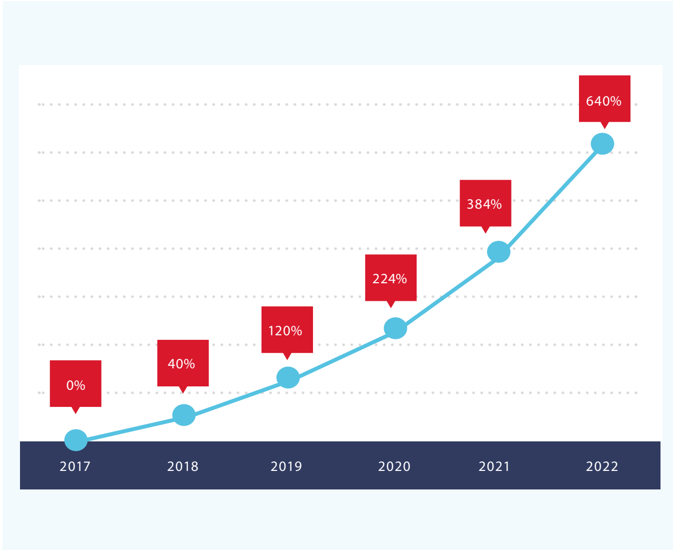

Figure 4: Projected Open Banking AIS daily query workload increase

Adding 640 percent to the existing query workload by year five is significant. However, the story is more complex. This workload will predominantly arrive in four discrete spikes (in four peak hours), driven by TPPs polling for updated data. When viewed by hour, as a percentage of the current total daily volume, Figure 5 shows the likely picture.

Figure 5: Day in the life - AIS query volumes by hour

Focussing on just these peak hours you get the growth curve in Figure 6.

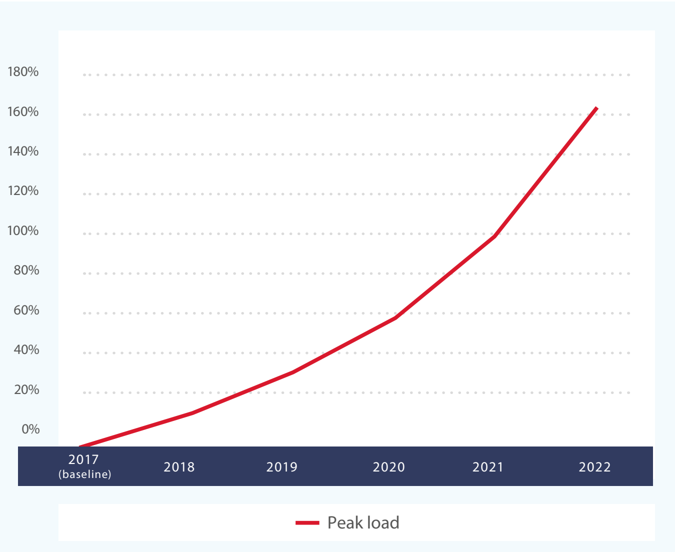

Figure 6: Peak hour workload as percentage of current daily workload

By year two, 30 percent of the current daily query workload is processed in peak hours four times a day. By year five, we are processing peak hours containing 160 percent of today's total daily workload (assuming an even distribution across that hour, which is optimistic). Compare this to a current peak hour which may process 10 percent of the day's workload.

3.3.2. Payment initiation

Taking industry statistics from April 2017[9], and comparing relative volumes for cards[10] and Faster Payments[11] shows us an order of magnitude difference between the volumes processed by card acquirers' existing payment schemes (see Table 1).

Data for April 2017	UK Faster Payments (SIP)	UK cards (debit & credit)
Transactions processed (000s)	79,998	1,386,000 (192,000 online)
Spend (£m)	66,000	58,000
Average transaction value (£)	825	42

Table 1: Payments statistics for April 2017

Looking back to 2016 industry data, scaling for 2017, then extrapolating a 5 percent move from card payments to Open Banking shows a step in transactions, as illustrated in Figure 7.

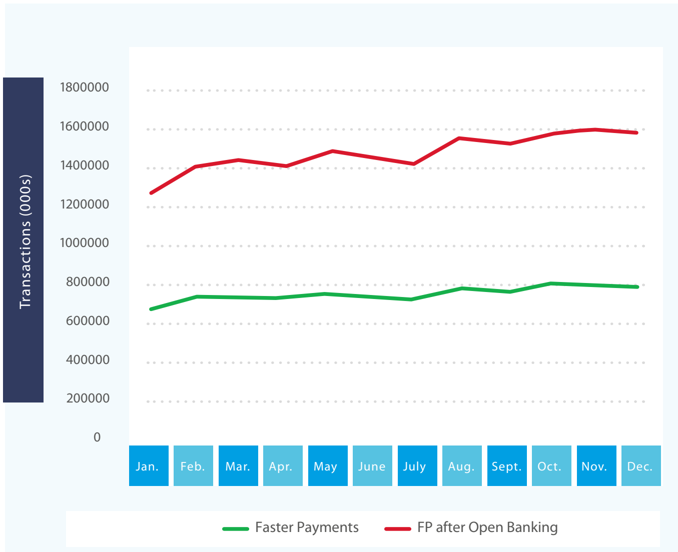

Figure 7: Growth in Faster Payments and impact of migration of cards traffic

Organic growth of this kind is almost certainly containable by all parties within the current scale of their payments estates. However, payments by card cope with much spikier workload for special events (such as charitable events). This workload is much harder to support on a platform executing a full Faster Payment.

A typical "day in the life" of Faster Payments infrastructure (including both Single Immediate Payments (SIP) and Standing Order Payments (SOP)) is shown in Figure 8.

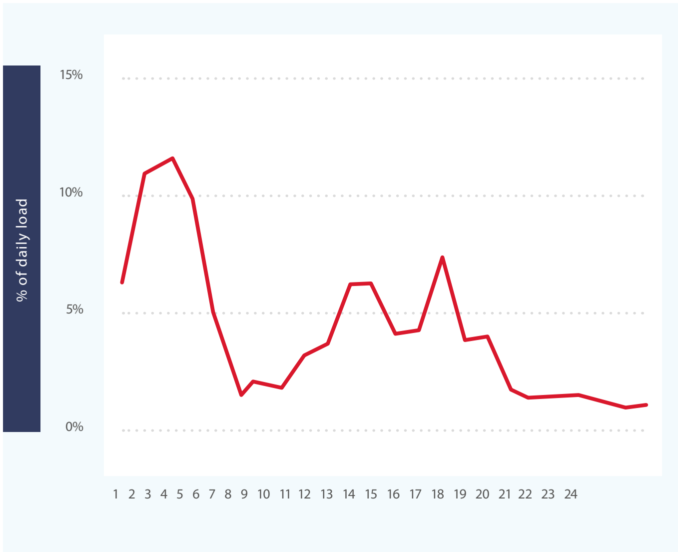

Figure 8: Typical Daily Profile for Faster Payments (SIP and SOP)

The key point to note is that the SOP peak at the start of the day generally represents the payment infrastructure running at maximum throughput.

If we overlay a two hour spike the same size as the rest of a day's processing (maybe 4 million payments across the entire scheme, which could be caused by some online event) we see the results shown in Figure 9.

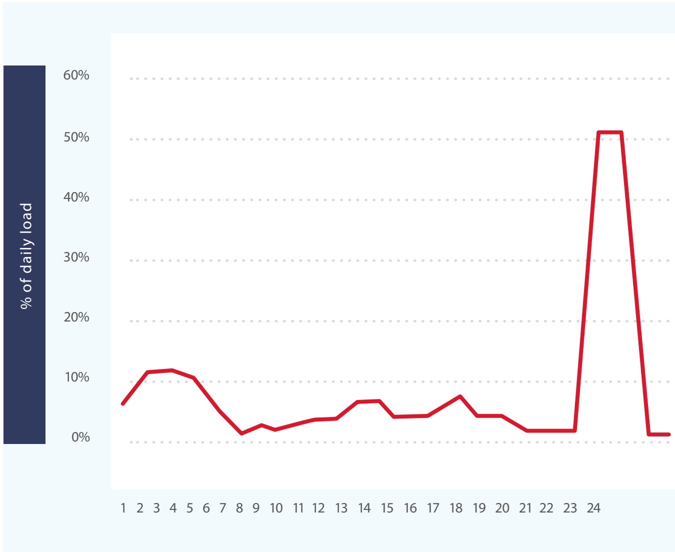

Figure 9: Impact an exceptional payments event has on on Payments workload

Payments infrastructures are not generally sized to cope with a peak throughput greater than 15 percent of their daily workload, let alone 50 percent. It is such peaky workloads that create a real threat to back-end infrastructures with the advent of PSD2 and Open Banking.

4. Areas of impact

The impact of Open Banking is pervasive, with each AIS or PIS request fanning out to create workload in each of the layers behind the channels. Concentrating on the primary components serving the requests (ignoring second order activities such as monitoring and management, fraud checks etc.) we can look at how the front-end volumes propagate to them. Figure 10 shows a logical topology for these key components.

For the two primary Account Information Service Provider (AISP) endpoints – balance enquiry and transaction query, and for a payment initiation, the assumed propagation to back-end components for a single payment is shown in Table 2. We assume that Locks and Blocks are checked for each transaction, and that Payment Initiation Service Providers (PISPs) will check the status of the payment after submission.

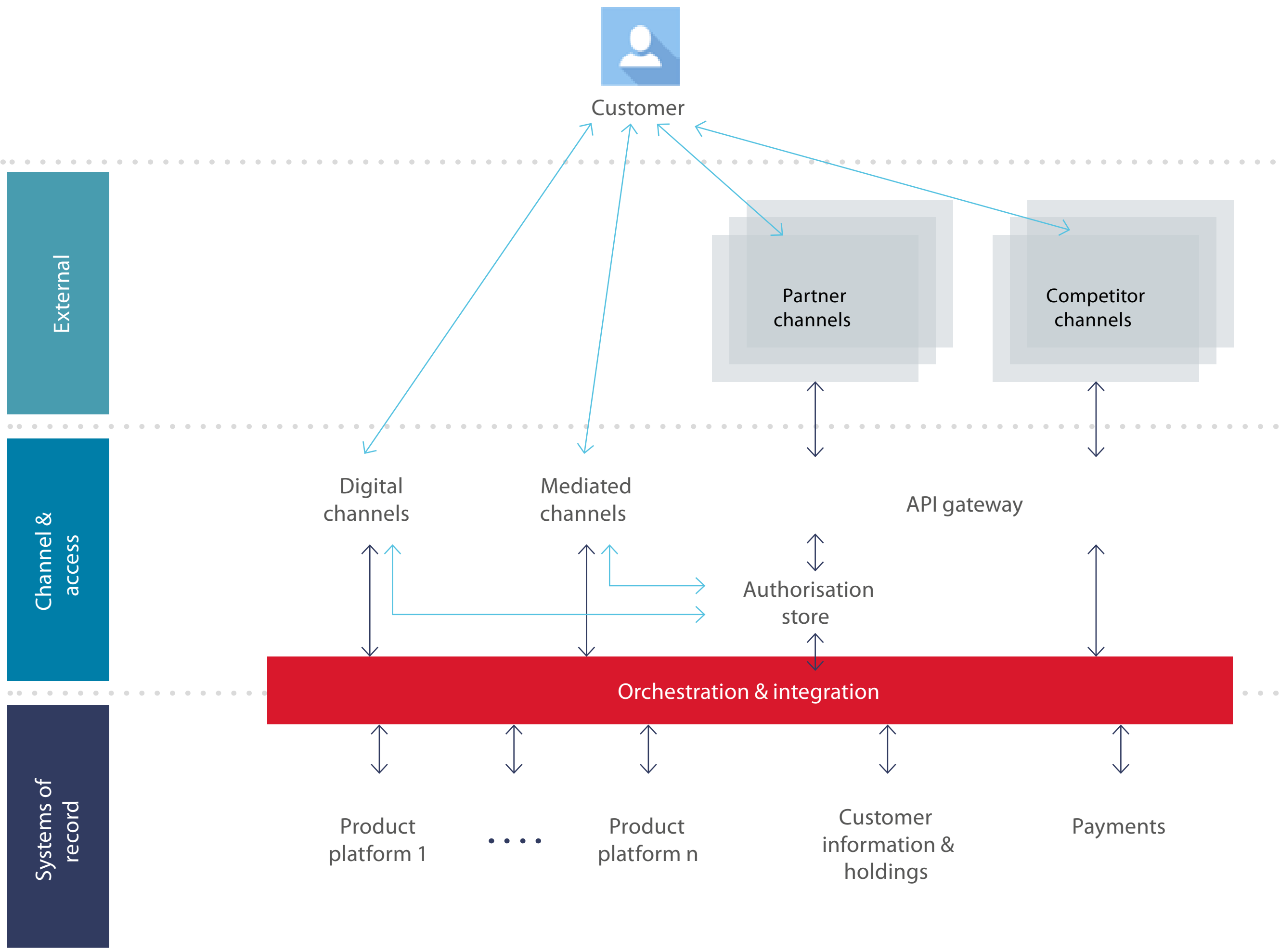

Figure 10: Unpacking the impacted components

Component	AISP balance query	AISP transactions query	PISP payment initiation
API gateway	1	1	3
Authorisation store	1	1	3 (R/W)
Orchestration	1	1	3
Customer information & holdings	2	2	2
Payments	0	0	1
Product/ accounting platform (e.g. current account)	2	2 (Complex)	3 (1 Read, 2 Write)
Totals	7	7	15

Table 2: Propagation of basic requests to underlying components

4.1. Impact of AISP workload

Taking the peak workload spikes discussed in section 3.3.1 (Account Information), and assuming a current unitary workload of 10,000 account queries per day (with a 10 percent daily peak), we get the peak workloads on back-end components shown in Table 3 and Figure 11.

Component	2017	2018	2019	2020	2021	2022
API gateway	2,000	2,400	6,400	11,600	19,600	32,400
Authorisation store	2,000	2,400	6,400	11,600	19,600	32,400
Orchestration	2,000	2,400	6,400	11,600	19,600	32,400
Customer information & holdings	4,000	4,800	12,800	23,200	39,200	64,800
Payments	-	-	-	-	-	-
Product/ accounting platform (e.g. current account)	4,000	4,800	12,800	23,200	39,200	64,800
Totals	16,017	18,818	46,819	83,220	139,221	228,822

Table 3: Peak transaction rates by key components for AISP workload

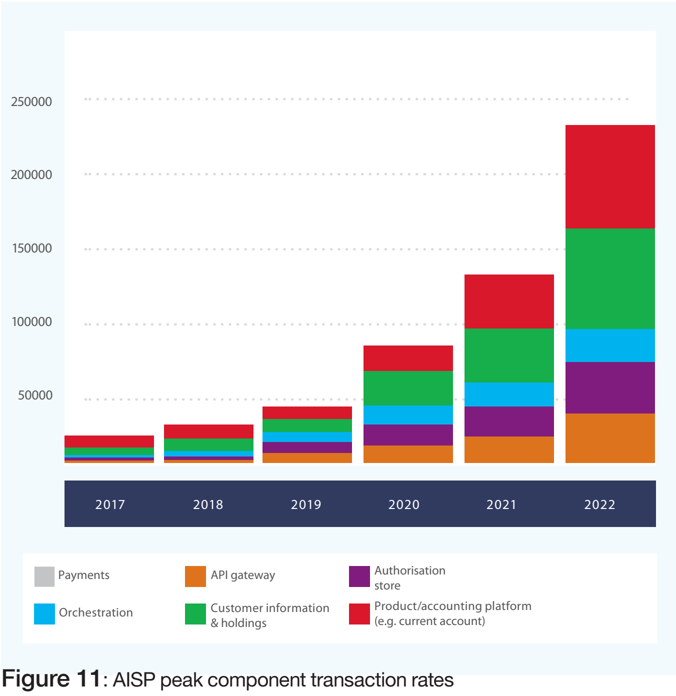

Figure 11: AISP peak component transaction rates

Increasing component workloads by more than an order of magnitude can lead to many complex failure modes when back-end systems become saturated. But, even at the most basic level, there is an escalating ratchet when back-end systems start to be saturated beyond their ability to respond.

Increasing latency

As systems start to reach their maximum capacity the apparent end-to-end latency of AISP calls into the bank will increase, as systems (usually one system in the end-to-end flow) starts to struggle. At workloads close to the maximum throughput of the most pressurised component, the small fluctuations in incoming load generally have sufficient jitter to allow processing to continue. However, response times will be highly variable, as the system uses short lulls in requests to catch up. In this phase some transactions may also time out – signalling incipient failure of the end-to-end processing chain. It is possible that increasing latency never manifests itself: with a sufficiently steep spike in workload the system will move immediately to unresponsive behaviour.

Unresponsiveness

Once the maximum throughput of a single component has been reached (receiving more requests per second than it is able to service), the end-to-end system will become completely unresponsive. The details of this failure mode may vary (with queue-fed components building up ever-increasing input queues, and API-based components failing to respond to requests), but the net effect is the same. While a trickle of transactions may still get through, the majority will stall and time out before responding.

Cascade failure

Components that rely on batch processing are particularly vulnerable to a failure cascade – especially if they are the component running close to their maximum throughput. Such systems typically suffer extreme performance degradation due to data contention during their batch windows. If a peak in query traffic coincides with such a window it can cause the component to stall in its normal processing – sending shockwaves to downstream processes and components (particularly components that are dependent on their data, but not technically linked. These components rely on accurate data being available after a certain time, since they assume that batch processing should be complete).

4.2 Impact of PISP workload

As covered in section 3.3.2, payments are particularly vulnerable to spikes in payment initiation. With a typical maximum transaction rate of 100 payments per second representing 15 percent of daily workload, a spike of 50 percent will yield the change in profile shown in Table 4 and Figure 12.

Component	Typical peak	Risk peak
API gateway	300	1,000
Authorisation store	300	1,000
Orchestration	300	1,000
Customer information & holdings	200	667
Payments	100	333
Product/accounting platform (e.g. current account)	300	1,000

Table 4: Peak transaction rates by key components for PISP workload

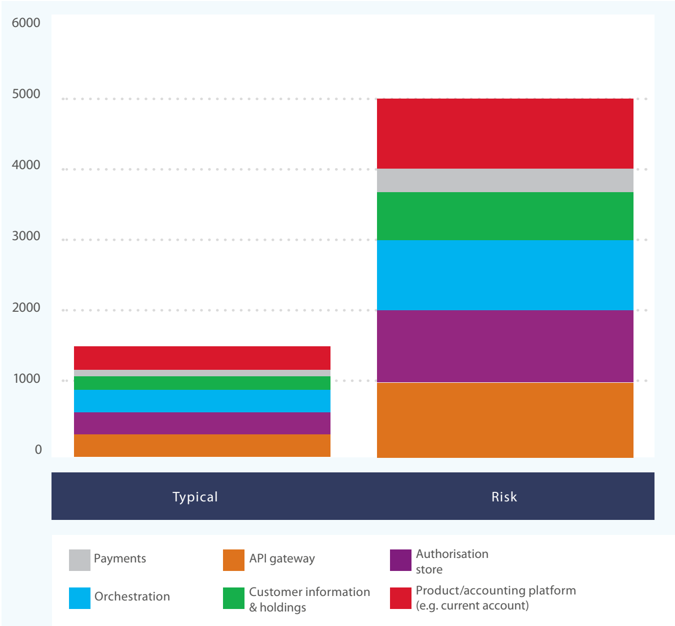

Figure 12: Payments peak component transaction rates

The failure mode for payments is rather more predictable than for AISP workloads, and can play out either "channel facing" or "scheme facing", or both.

Scheme facing

Saturation of components facing the Faster Payments scheme, in the best case, leads to having to respond with a QA to the scheme. At processing rates that overwhelm the gateway this could result in becoming unresponsive to the scheme (or having to temporarily withdraw). But, in most infrastructures, this would be a very extreme case, as upstream components should generally be capable of handling extreme workloads).

Channel facing

For payments initiated by the channels (and likely the PISP Open Banking API), failure modes are similar to the AISP scenarios – with increasing latency on responses giving way to unresponsiveness (with the same potential for downstream cascade failure). The multi-step nature of payment initiation for Open Banking (Setup, Authorise, Submit, Query Status) also means that transactions could fail at any one of a number of steps – causing retries from the client, which hasten incipient failures by artificially inflating the transaction rate.

Downstream failures

Failure of one of the other scheme members, or the scheme itself is another potential risk at extreme volumes. This would drive additional scheme side processing at a time when the channel-facing side is also likely under pressure. It is sensible to be cognisant of this risk when designing mitigation strategies for overall workload (to ensure it is not capable of acting as the trigger for a failure cascade internally), but not to over-plan, since it is ultimately an externality to the bank.

5. Mitigation strategies

The "simple" emergency mitigation strategy of turning off the problem at source by switching off servicing of the AISP and/or PISP APIs is only an option in the direst of circumstances. The scenarios described above will become the "new normal", leaving some possible mitigation strategies.

5.1. Additional resources

To cope with additional volume, the most straightforward approach is to provide additional resources (CPU and/or memory) to the components under pressure. This is only possible if the component does not encounter an architectural bottleneck (usually in the data access layer or an existing single threaded process that cannot use additional resource threads). When this approach is tractable, returns are usually broadly linear, e.g. double the CPU and you double the throughput. With the Open Banking workload growth projected to accelerate in an exponential fashion, this can become unaffordable in the long run.

5.2. Throttling and traffic shaping

It is advisable to provide a means of throttling account information requests and to shape the traffic to remove transient peaks. This capability is required: at the API tier, to protect the Open Banking infrastructure, and at the orchestration and integration tier, to protect existing systems of record. The latter serves to flatten peak loads and can also be used to reduce traffic during known periods when systems are heavily loaded, e.g. during batch windows.

However, great care is required when throttling starts to delay responses times, as eventually it can result in time-outs. The normal strategy of retrying a service call after a time-out will increase calls, consuming more resources and potentially producing cascading failures. To give systems the ability to recover and respond, throttling must be combined with a circuit breaker pattern.

5.3. "Cloudification"

In many ways an extension of "additional resource", moving applications to cloud-based infrastructure provides far more elastic management of workload. However, it is not a magic bullet, as moving an application to cloud is not simple. The application may need to be re-structured to remove architectural bottlenecks (such as single threads that hold critical locks or resources).

It will never be practical to move all applications to the cloud, as their underlying architecture is often too wedded to specific deployment models. This is particularly challenging for applications built using traditional techniques for high availability and disaster recovery, such as active/passive configurations with replicated databases.

5.4. Data fabric

The best mitigation for account information workload is shortening the path to the data (reducing the number of components impacted by the volumes, while decoupling them from the back end) and introducing a data component specifically designed to service queries up to extreme rates. This component should have all changes to the systems of record pushed to it to decouple them from client queries.

As soon as the query workload from Open Banking exceeds the workload of updating the systems of record, then it is lower cost to push the changes through to a central point of truth optimised to handle them, and have it service them directly in the orchestration/integration tier. The key success factor is to ensure that the new access path replaces all links in the chain that could become saturated.

5.5. Qualified Accept at the channel for payments

The synchronous nature of Faster Payments, and tight round trip timing requirements make the process fragile in the face of arbitrary workload. To be certain you can cope with such workloads when back-end components may become unresponsive, you must provide an equivalent to a QA to the channel (API), as well as providing a QA capability to the Payment Scheme. This needs to be coupled with an elastic approach to queueing payments upstream, and throttling injection to avoid exacerbating problems downstream.

5.6. Re-architect or re-platform

As a last resort, where a component can simply never meet expected demand, then either re-factoring it to remove limiting bottlenecks or replacing it are the only remaining options. The former is complex and usually involves reliance on very scarce knowledge, while starving the business of change capacity. The latter creates a migration backlog, again at the expense of business change capacity.

6. How IBM can support a structured mitigation approach

6.1. Performance engineering, modelling and early technical proving

Performance Engineering aims to reduce risk, and identify and resolve any performance and scalability issues early in the solution lifecycle. It is an established discipline for optimisation of design and corresponding runtime environments of IT systems. The IBM® Performance Engineering & Management Method helps to ensure that performance is a prime consideration throughout the solution lifecycle.

Performance modelling is used to transform expected business volumes into expected transaction rates, enabling the identification of potential bottlenecks in multi-tiered, distributed systems. It is used to identify areas of risk in the overall system landscape that require further treatment. It may be possible to precisely identify the mitigation required, but, in many cases, there will be insufficient data to understand how system components are likely to respond under increased load.

Technical proving focuses on these high-risk areas, verifying whether a component or set of components will meet performance and availability requirements. By creating prototypes to stress-test system components, technical proving can support calibration of performance models and accurate predictions of system component behaviour under increased load. It is particularly useful to identify areas where it will not be sufficient to increase systems resources, and instead architecture and design refactoring will be required.

6.2. End-to-end performance testing and resource tuning

Performance engineering and modelling can only provide insight into system behaviour up to a point, after which end-to-end testing is required. This is because there is considerable complexity in the interaction between the components of modern, multi-tiered, distributed systems. Increasing the resources in one area can cause bottlenecks and failures elsewhere. Similarly, one component responding poorly can cause catastrophic, cascading failures in other areas.

To be effective, end-to-end performance testing requires highly skilled test professionals to design the end-to-end tests, test tooling to drive the volumes through the system, monitoring of all components to identify bottlenecks and most importantly, representative test environments. The challenges of creating representative end-to-end test environments are notorious and complex, requiring tools to manufacture test data, heterogeneous platforms, and the ability to create representative test stubs, access test systems and run daily batch schedules faster than real time while maintaining synchronicity of test data.

IBM IGNITE leverages an optimisation engine that offers complete integration test automation (web, mobile, API test automation), prevents defects with cognitive defect analysis and combines seamlessly with test data and test management. This capability helps clients address market demands for faster testing, reduction of cost, increased automation, decrease of test inventory and fewer defects.

6.3. Architecture mitigation, including multi-threading, circuit breakers and throttling

Where technical proving has demonstrated that a component can never meet the expected demand, it must be re-factored to remove limiting bottlenecks, or replaced. Microservices patterns provide an agile approach to refactoring legacy applications so they become fully cloud-ready (read on to learn more about IBM Cloud Innovate).

At the macro level, throttling account information requests and shaping traffic to remove transient peaks is advisable. At the API tier, this capability can protect Open Banking infrastructure, while at the orchestration and integration tier it can safeguard existing systems of record. It is important to combine throttling with appropriate time-out settings and use of the circuit breaker pattern, as this protects components higher up the architecture stack and provides systems of record with the slack they need to recover and respond.

IBM DataPower® Gateway is an enterprise-grade, proven, security-rich API gateway available in multiple form factors. Offering high performance and scalability, it provides built-in policies for security, traffic management and mediation.

6.4. Cloud migration

Migrating systems to the cloud can provide greater scalability and affordability at scale. It is unlikely that systems of record will have been architected with cloud in mind, so the benefits are probably limited to enhanced scalability of the underlying cloud platform and reduced cost of compute resources. To achieve the full elasticity of scale promised by cloud, you must redevelop applications as cloud-native microservices.

The IBM Cloud Innovate method helps clients with this journey by providing reliable, cost-competitive, repeatable services for migration, modernisation, operation and management of IT portfolios. Specifically, it will enable you to transform enterprise application portfolios by migrating to cloud, modernise and transform existing applications using microservices and build new, cloud-native applications.

6.5. Data fabric implementation

In a Hybrid Cloud world, managing data resources, which usually reside in physical silos, in an agile, platform-independent manner is a major challenge. Applications are generally stateless and can be provisioned quickly in various environments. Data is stateful and dependent upon underlying physical data stores. Data fabric is an IT term used to describe data that is accessible where and when needed, not reliant upon the location of, and persistent connectivity to, physical data stores.

Data fabric describes a distributed, memory-based data management platform which uses cluster-wide resources (memory, CPU, network bandwidth and optionally local disk) to manage data. You should consider data fabric as an architectural concept rather than a product. Its primary features are continuous availability, very high performance, portability and linear scalability for data-intensive applications without compromising on data consistency. To address non-functional concerns with regards to Open Banking, IBM advocates the construction of a data fabric layer alongside existing middleware to support Open Banking APIs.

The IBM data fabric solution is based on IBM WebSphere® eXtreme Scale (WXS) - an elastic, scalable, in-memory data grid which dynamically caches, partitions, replicates and manages application data across multiple servers. Integrated with this high-performance data caching capability is IBM Db2® Advanced Enterprise Server Edition, providing a persistent Operational Data Store (ODS) to ensure data consistency and availability – e.g. rapid recreation of cache in the event of a severe outage.

To populate the IBM Data Fabric without repeatedly querying systems of record to “pull” information, IBM InfoSphere® Change Data Capture (CDC) offers a low-latency solution, which “pushes” systems of record updates as they occur to a messaging layer (message queuing or direct Transmission Control Protocol/Internet Protocol), for delivery to the Db2 and WebSphere eXtreme Scale systems in near real time.

6.6. Payments estate renovation

Payments estates have grown up organically, with even newcomer schemes such as UK Faster Payments pre-dating the explosion in consumer-driven demand. To cope with the unprecedented peak workloads that may be generated by PISP use cases as workload migrates from cards (particularly as friction around authentication is reduced over time), payments estates need elasticity to smooth peak workloads, and circuit breakers to intervene in the face of overwhelming workload. To enable the renovation and progressive upgrade of the payments estate, IBM Financial Transaction Manager (FTM) provides a highly scalable and robust payments hub capable of smoothing traffic and deploying circuit breaker patterns.

7. Conclusion

Arguably, the volume estimates in this document are overly pessimistic and illustrate a worst-case scenario. However, failure to mitigate the impact of peak loads will lead to failure of back-end systems. Such failures will be immediately visible to customers and cause major reputational damage; the last thing the banks need when coping with regulations designed to limit their control within the market. Impact will be direct to the bottom line, as customers move to more responsive banks. Furthermore, this is only the first wave of regulation: other financial products will soon need to be opened up too.

Banks must act now to protect their critical systems of record from this new and growing front-end demand. IBM can help overcome these threats by working with you to create the predictive performance and capacity models that will help to identify potential bottlenecks. This will help us and you to identify areas of risk that requires technical proving, end-to-end performance testing and where required performance tuning, architecture remediation or design change. We can help you migrate to the cloud, which together with a data fabric deployment have the potential of achieving elastic scalability. We have the experience of deploying IBM Financial Transaction Manager to achieve extreme levels of scalability and resilience.

About the authors

Pieter Lindeque is the Chief Technology Officer for IBM Global Business Services in the United Kingdom and Ireland. He is an IBM Distinguished Engineer and certified Executive IT Architect who consults to CIOs and CTOs on digital transformation, integration and resilience. His clients include some of the largest financial services and public sector organisations, where he has shaped and led the architecture for complex systems integration programmes. He is known for his ability to enable implementation at scale. Pieter can be reached at lindeqp@uk.ibm.com.

Paul Lucas is a certified Executive IT Architect specialising in Open Banking, Payments Systems and large-scale systems integration. He has considerable experience in systems integration, application architecture, and software design and development. In recent years he has focused on integration and simplification in the payments systems space. He has led major transformation programmes for both commercial and retail banks from strategy through to implementation. Paul can be reached at paul_lucas@uk.ibm.com.

Footnotes

1. Systems of record provide the authoritative source for business data. The key system of record for Open Banking is the ledger/accounting platform for the customer's current account, providing balance and transaction information.

2. Barclays, Nationwide, Bank of Ireland, Danske Bank, HSBC, RBS, Lloyds, Santander, AIB.

3. www.openbanking.org.uk

4. www.openbanking.org.uk/open-data-apis

5. www.openbanking.org.uk/read-write-apis/account-transaction-api/v1-1-0

6. www.openbanking.org.uk/read-write-apis/payment-initiation-api/v1-1-0

7. There are other emerging standards from the Berlin Group and STET. But, at the time of writing, they lack the level of detail required for unambiguous implementation.

8. en.wikipedia.org/wiki/Representational_state_transfer

9. Representative of both schemes, which both exhibit steady growth.

10. Including both debit and credit card transactions.
Source: www.theukcardsassociation.org.uk/wm_documents/Card%20Expenditure%20Report%20-%20April%202017.pdf

11. Focussing only on Single Immediate Payments (SIPS).
Source: www.fasterpayments.org.uk/statistics

IBM United Kingdom Limited
PO Box 41, North Harbour
Portsmouth
Hampshire, PO6 3AU

Produced in the United Kingdom
March 2018

89014689USEN-00