CRIF DIGITAL

The golden age of data in financial services

June 2021

01
Executive summary

Financial services is a data-intensive industry in which banks and insurance companies have historically held large quantities of information.

The development of the internet, the increase in new channels, and the introduction of technologies including cookies and trackers have helped not only to create a better user experience, but also to exponentially increase the amount of collected data, for the most part dark data. With the availability of new digital technologies and a higher degree of interconnectivity and interactions, new data sources have emerged that are a potential treasure trove for businesses.

Up to the early 2000s, organisations held their data pools within the boundaries of their own technology infrastructure. During the dotcom era, PC accounting software emerged that, as one of its many applications, started to allow the screen scraping of account transaction information as input into personal tax calculations.

Designed to force competition in the banking market by making the banking data available to new types of financial services providers, Open Banking initiatives were initiated by several regulators, creating new opportunities for value propositions and better serving customers based on their data. Whilst the uptake is slow in some markets due to the need to develop consumer confidence in data sharing, Open Banking is already evolving into Open Finance as the roll-out of APIs and the availability of data is extended to additional products and services.

CRIF Digital shows that the value to financial services organisations, and to their customers, includes an increased redemption rate, better risk management and risk profile quality of their lending portfolio, to name just a few. These are proof of the potential value held in data.

Drawing on our 32 years' experience in insights and analytics, there are a few valuable lessons learned that should be considered when looking at deriving value from data:

01 **The importance of the availability, quality and relevance of data** – without heterogeneous data to start from, it is not possible to speed up the development of all the analytics necessary to leverage this new information "gold mine". The pre-processing of data is crucial to making unstructured data easier for machines to analyse and to increasing data quality.

02 **The value of data increases as the data available for processing increases** – Open Banking is currently the most talked about opportunity for deriving value from data. From the experience gained through CRIF Digital, data sources rarely substitute each other, but in almost all cases they complement each other, greatly improving the value of the analytics.

For example, CRIF Digital has shown this through the combination of Open Banking data with its credit bureau information. It is unstructured data that makes it possible to enrich the "traditional" data source with new information – e.g. knowing if an account holder is going into overdraft as a result of losing his/her job (no more salary transactions) or because he/she has simply started to spend more and what he/she is spending it on. Combining structured and unstructured information guarantees the best analytics independently of the data assets available in different geographies.

03 **Ultimately, it is all about value to the end user** – consumers and businesses expect a seamless digital experience. It is all about value to users and how the analysis of customer data can improve their financial lives.

Introduction: the growing number and size of data pools

With the availability of new digital technologies and communication systems, the higher degree of interconnectivity and interactions, the growing sophistication of web technologies, social media, mobile computing and online marketing, new data sources have emerged that are a potential treasure trove for businesses.

The variety of accessible data presents an unexplored value for businesses when it comes to identifying challenges and pursuing new opportunities. Customer insights become richer and more valuable as the availability of structured and unstructured customer data grows, which are significantly different in nature. Until recently, most analytical capabilities have focused on the latter.
But how do these data types differ?

"Structured data uses a predefined and expected format. This can come from many different sources, but the common factor is that the fields are fixed, as is the way that it is stored. This predetermined data model enables easy entry, querying, and analysis."[1]

Structured data is easily available as a record of an event. "Because each field has a defined purpose, it makes it easy to manually query..... and for machine learning algorithms to identify patterns."

[1] https://blogs.oracle.com/bigdata/structured-vs-unstructured-data

Structured data follows a predefined schema, whereas unstructured data is the opposite. "Unstructured data can come in all shapes and sizes. Though typically text, unstructured data can come in many forms to be stored as objects: images, audio, video, document files, and other file formats."

Unstructured data originates from new technologies and communication platforms and cannot be analysed using a predefined data model and controlled navigation is not possible as its content or structure does not follow a pattern. "Social posts are a specific example of unstructured data. The metrics behind each social media post—likes, shares, views, hashtags, and so on—are structured, in that they are predefined and purposeful for each post. The actual posts, though, are unstructured."

For data analytics and data science purposes, the datasets used can be described and categorised by five criteria, as illustrated in Fig. 1. The first four categories are some of those most frequently quoted in academic literature and, based on the experience from CRIF Digital, it is also important to specify a fifth category: certified data. This refers to data that has been processed through structured quality checks ensuring it meets set quality criteria related to an intended use. Certified data is typical of credit bureaus and other official data sources that has been cleaned, screened and validated within parameters set in data governance structures.
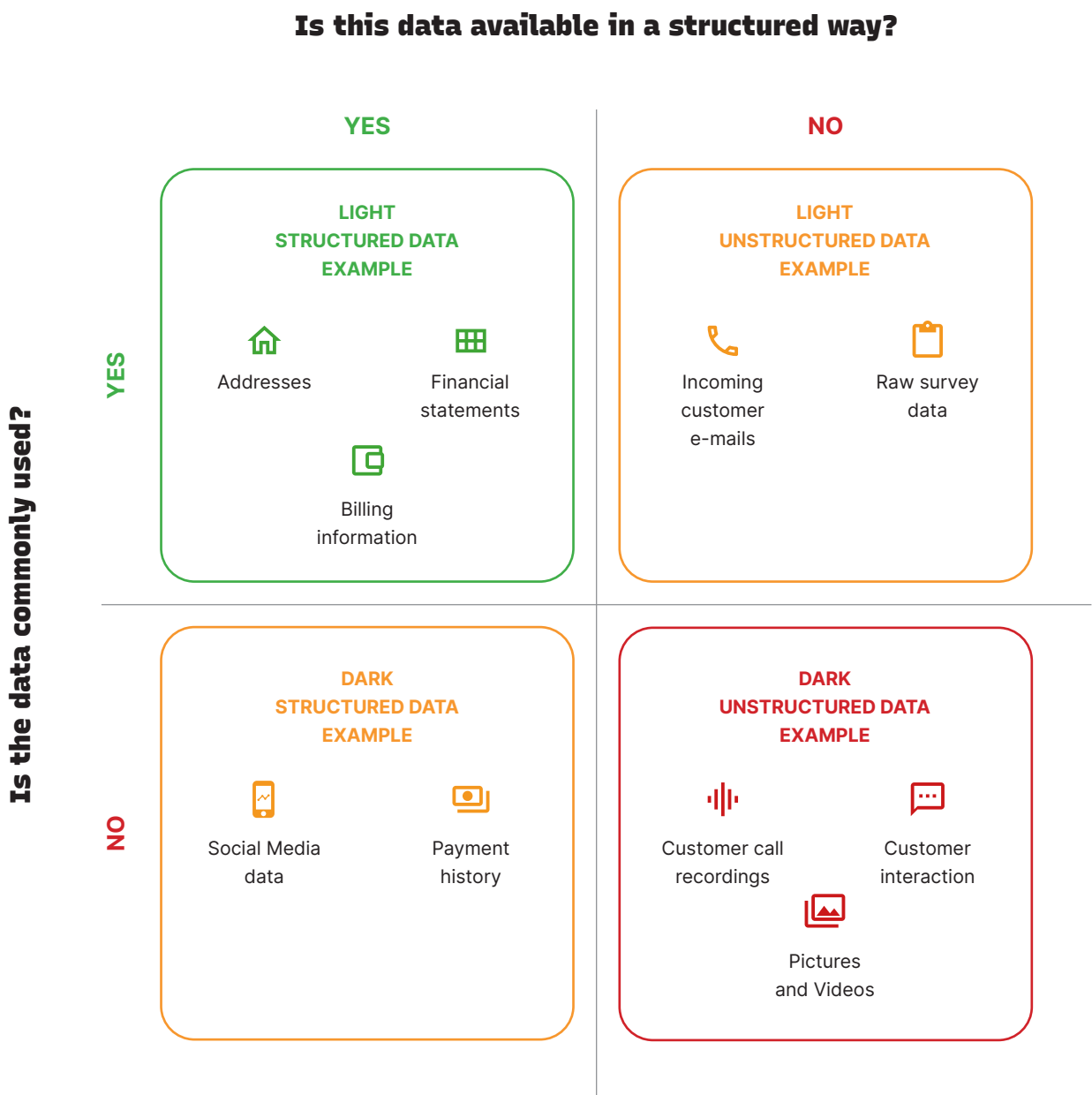
Main data categories[2]
Fig. 1

DEFINING CRITERIA	DATA TYPES	
Consistency & Organisation	**Structured** (relational database tables following a pre-defined schema)	**Unstructured** (character and binary data with non-identifiable structure; cannot be stored in rows and columns in a relational database)
Frequency of Use	**Light** (commonly used)	**Dark** (rarely used)
Extent/focus	**Small** (schema-dependent dataset, which is difficult to scale once it contains content)	**Big** (large amounts of data that is generated automatically by business/customer activities and that are analysed in search for unspecified insights to be discovered)
Gathering Approach	**Deep** (gathered in large quantities, systematically on daily basis)	**Targeted ad hoc** (targeted gathering of certain datasets for specific purposes)
Quality and attribution	**Certified** (set or exceeding data quality guaranteed through dedicated processing)	**Uncertified** (data whose quality is not guaranteed and that has not been qualitatively screened or processed)

[2] Sint, Rolf & Stroka, Stephanie & Schaffert, Sebastian & Ferstl, Roland. (2009). Combining Unstructured, Fully Structured and Semi-Structured Information in Semantic Wikis.

Example of possible permutations within the data categories

Fig. 2

The five criteria can be combined into several possible permutations that describe how datasets are collected, organised, and used for different apurposes. Examples within two of these data categories are illustrated in Fig. 2.

Is this data available in a structured way?

YES   **NO**

Is the data commonly used?

YES

**LIGHT
STRUCTURED DATA
EXAMPLE**

Addresses   Financial
statements

Billing
information

**LIGHT
UNSTRUCTURED DATA
EXAMPLE**

Incoming
customer
e-mails   Raw survey
data

NO

**DARK
STRUCTURED DATA
EXAMPLE**

Social Media
data   Payment
history

**DARK
UNSTRUCTURED DATA
EXAMPLE**

Customer call
recordings   Customer
interaction

Pictures
and Videos

It is interesting to note how the applicability of the category descriptors can change over time. Before the rollout of Open Banking in any specific market, even transaction information could have been classified as dark data as it was occasionally used through screen scraping[3] by providers of the earliest Personal Financial Management (PFM)[4] propositions that were rolled out immediately after the dotcom boom.

This data has moved from being defined as Dark to becoming Light as Open Banking regulations are driving the development of a wave of value propositions that are leveraging transaction information. While there are different ways to describe and categorise data types, they share a common characteristic: they are growing exponentially as shown in Fig. 3. In particular, machine-generated unstructured data has grown exponentially due to the fast-growing digital lifestyle of customers[5], with unstructured data estimated to be 85% of all existing data[6].

[3] Process of collecting screen display data from an application. The data is translated so that another application can display it. In the past, screen scraping allowed third-party companies to access financial transaction data by logging into digital portals on behalf of a financial institution's customers.

[4] Allows users to aggregate financial transactions in one place and use that data to manage their money. PFM was first introduced by Intuit and Microsoft, paving the way for online PFM software.

[5] Eberendu, Adanma Celia (2016). Unstructured Data: an overview of the data of Big Data - International Journal of Computer Trends and Technology (IJCTT) V38(1): 46-50, August 2016.

[6] De Boe, B. (2014). Use Cases for Unstructured Data – InterSystems White Paper, InterSystems Corporation.

60,000

Growth of data
Fig. 3

50,000

Extracting value from the increasing amount of data has been limited by the availability and affordability of processing and storage technology, but this is no longer the case. Cloud technologies and artificial intelligence enable the storage and analysis of both structured and unstructured data and allow companies to determine customer reactions, identify product preferences, personalise products or detect fraud in real time, etc. Overall, this is a wealth of insights and information that is waiting to be fully exploited by the financial services industry.

40,000

Exabytes (Billions of GB)
■ Structured data
■ Unstructured data

30,000

20,000

10,000

0

2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020

03
Open banking and open data

Financial services is a data-intensive industry in which banks and insurance companies have historically held large quantities of information.

This data falls into the categories illustrated above with a large proportion of dark data whose potential value has been mostly left unleveraged. The development of the internet, the increase in new channels, and the introduction of technologies including cookies and trackers have helped not only to create a better user experience, but also to exponentially increase the amount of collected data, for the most part dark data.

Up to the early 2000s, organisations held their data pools within the boundaries of their own technology infrastructure. With the development of digital channels and interfaces, information about user experience behaviour, transactions, balances, etc. has been exposed through software interfaces.

During this time, screen scraping was developed as a technology to capture information that other organisations were holding on behalf of their customers. During the dotcom era, PC accounting software emerged that, as one of its many applications, started to allow the screen scraping of bank account and credit card transaction information as input into personal tax calculations.

Account aggregation information solutions were launched in the early 2000s followed by short-term lending and other aggregation- based propositions that brought the value of data-driven account insights to mainstream attention.

The emergence of Application Programming Interface (API) technologies has made screen scraping obsolete, bringing a much more cost-effective and secure way to capture or exchange data. The possibility for permissioned third parties to access data through APIs has become a feature of the entire market with Open Banking.

The EU Payment Services Directive 2 (PSD2)[7] introduced provisions for specific transaction information and payment initiation services available to licensed and permissioned third parties. Designed to force competition in the banking market, Open Banking initiatives were introduced by several other regulators with different approaches (prescriptive and industry-driven), scope (products and datasets) and reach (applied to the whole market or only to larger organisations).

Additionally, in some countries, Open Banking-like initiatives and data access through APIs has been initiated by banks. Fig. 4 provides a snapshot of the status of Open Banking initiatives globally[8], while Fig. 5 shows the timing of Open Banking planning and rollout in different countries.
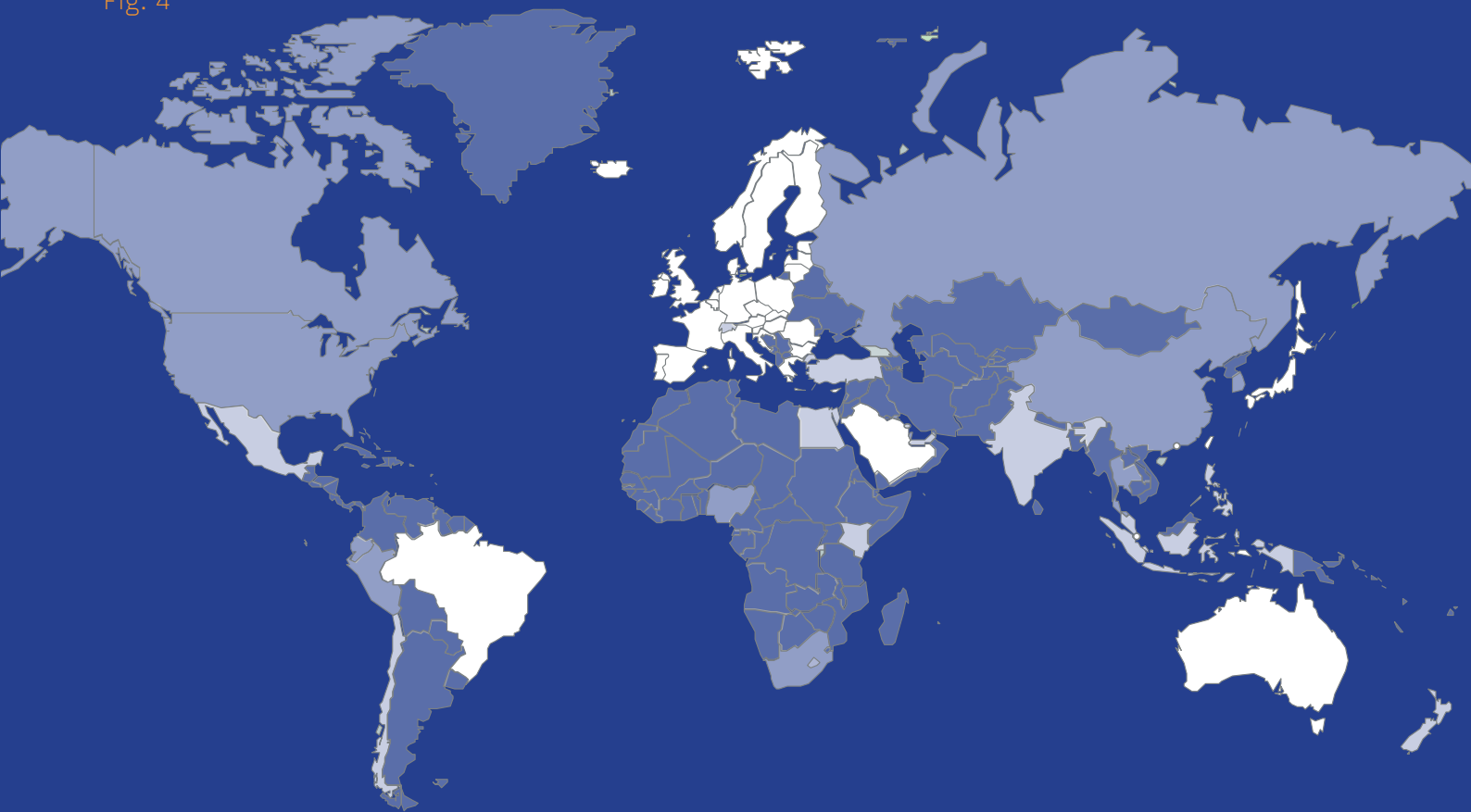
[7] https://ec.europa.eu/info/law/payment-services-psd-2-directive-eu-2015-2366_en

[8] Various sources. CRIF analysis

Status of Open Banking initiatives around the globe

(as of June 2021)

Fig. 4

- No reported activity
- Evaluation
- Regulation in development
- Roll-out
- Fully Operational

Roll-out

Australia	Lithuania
Austria	Mexico
Belgium	Norway
Brazil	Netherlands
Bulgaria	Poland
Estonia	Romania
France	Saudi Arabia
Germany	Slovenia
Greece	Spain
Italy	Sweden
Japan	UK
Latvia	

Regulation in development

Brunei
Burma
Chile
India
Kenya
Malaysia
Papua New Guinea

Evaluation

China
Ecuador
Nigeria
Peru
Russia
Vietnam

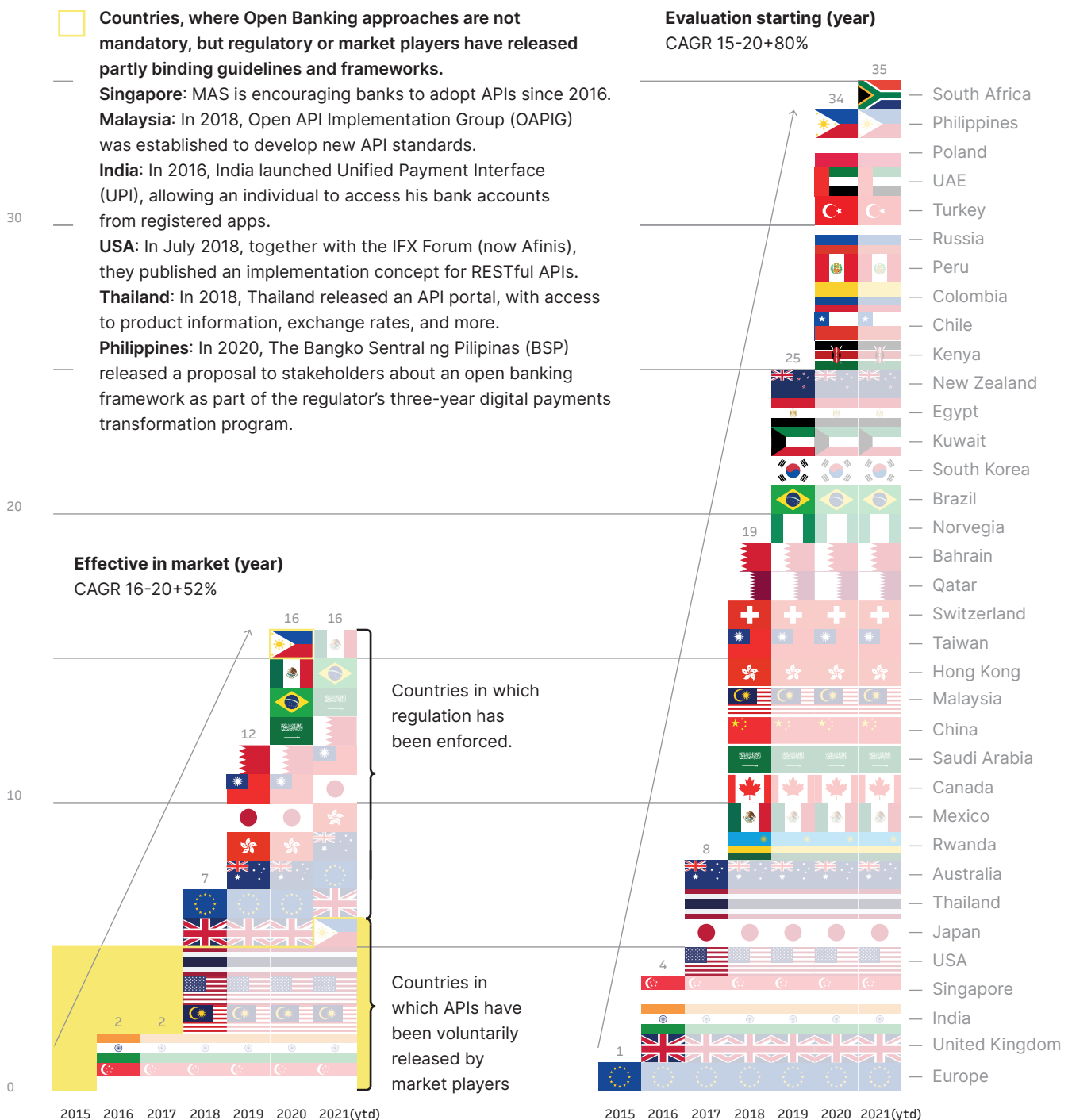**No reported activity**

Rest of the World

Fully Operational

None

Evolution of Open Banking initiatives around the globe
(evaluation and implementation by country - status as of June 2021)

Fig. 5

Countries, where Open Banking approaches are not mandatory, but regulatory or market players have released partly binding guidelines and frameworks.

Singapore: MAS is encouraging banks to adopt APIs since 2016.

Malaysia: In 2018, Open API Implementation Group (OAPIG) was established to develop new API standards.

India: In 2016, India launched Unified Payment Interface (UPI), allowing an individual to access his bank accounts from registered apps.

USA: In July 2018, together with the IFX Forum (now Afinis), they published an implementation concept for RESTful APIs.

Thailand: In 2018, Thailand released an API portal, with access to product information, exchange rates, and more.

Philippines: In 2020, The Bangko Sentral ng Pilipinas (BSP) released a proposal to stakeholders about an open banking framework as part of the regulator's three-year digital payments transformation program.

Evaluation starting (year)
CAGR 15-20+80%

Effective in market (year)
CAGR 16-20+52%

Countries in which regulation has been enforced.

Countries in which APIs have been voluntarily released by market players

The scope of Open Banking, initially starting with payment account access, has been broadened to a wider range of financial services products. Open Finance operates - where available - on the same principles as Open Banking, with access based on permissions and licenses. This extends to savings, insurance, mortgages, investments, pensions, and consumer credit, providing consumers and businesses with more control over a wider range of their financial data. From a data perspective, this significantly increases the amount of data available for processing, even if customers have reservations about the use of their information. "What we say and what we do when it comes to sharing personal data, notably around Open Banking, are two distinct things"[9] according to an article published by ING on the back of their research on consumer attitudes towards data sharing "ING survey: We're still suspicious about Open Banking".[10]

"The findings show that sharing personal data and interacting with technology in new ways are not developments that people unanimously support"[11] with all the limitations to adoption and development deriving from consumers not being overly confident about sharing their data with financial institutions or authorised third parties. The term Open Banking refers to transaction information from current accounts, while many countries and jurisdictions have broadened the reach of the permission and licensed-based exchange of information to other products, such as mortgages, insurance and investments. For example, Indonesian regulations primarily cover payment information (Open Banking), while in Australia the scope includes current accounts, loans, mortgages and credit cards, with future extension to utilities and telecom data (Open Finance). The UK and the European Union have recently started the process of broadening their Open Banking regulations to Open Finance through the inclusion of additional financial product information. Fig. 6 provides an illustration of the main characteristics of Open Banking and Open Finance around the world.

---

[9] https://think.ing.com/articles/what-we-say-and-what-we-do-differ-in-a-tech-world

[10] Ibid.

[11] Ibid.

Status of Open Banking and Open Finance around the world

Fig 6

DRIVING FORCE	REG. APPROACH	TARGET	TYPE/SCOPE
Regulatory led	Prescriptive	Focused (specific scope)	Not Finalised — Open Banking — Open Finance
Market led	Industry-driven	Wide (industry-wide)	

North and South America

Country	Driving Force	Reg. Approach	Target	Type/Scope
Canada	(NOT YET SPECIFIED)			
USA	Market led	Prescriptive	Wide	
Mexico	Regulatory led	Industry-driven	Wide	
Colombia	Market led	Industry-driven	Wide	
Peru	Market led	Industry-driven	N/A	
Brazil	Regulatory led	Prescriptive	Focused	
Chile	Market led	Industry-driven	Wide	

Europe, CIS and Africa

Country	Driving Force	Reg. Approach	Target	Type/Scope
UK	Regulatory led	Prescriptive	Focused	
EU	Regulatory led	Prescriptive	Focused	
Switzerland	Market led	Industry-driven	Wide	
Russia	Regulatory led	Industry-driven	Focused	
Turkey	Regulatory led	Industry-driven	N/A	
Egypt	Regulatory led	Industry-driven	Wide	
Nigeria	Regulatory/Market led	Industry-driven	Wide	
Kenya	Regulatory/Market led	Industry-driven	Wide	
Rwanda	Regulatory led	Industry-driven	Wide	
South Africa	(NOT YET SPECIFIED)			

Middle East and Central Asia

Country	Driving Force	Reg. Approach	Target	Type/Scope
Kuwait	Market led	Industry-driven	N/A	
Bahrain	Regulatory led	Industry-driven	Wide	
Qatar	Market led	Industry-driven	N/A	
UAE	Regulatory/Market led	N/A	N/A	
Saudi	Regulatory led	Industry-driven	Focused	
India	Regulatory/Market led	Industry-driven	Focused	
China	Market led	Industry-driven	Focused	
Hong Kong	Regulatory led	Industry-driven	Wide	

South East Asia, Far East and Asia Pacific

Country	Driving Force	Reg. Approach	Target	Type/Scope
South Korea	Regulatory led	Industry-driven	Wide	
Japan	(NOT YET SPECIFIED)			
Singapore	Market led	Industry-driven	Wide	
Thailand	Regulatory/Market led	Industry-driven	Focused	
Malaysia	Regulatory/Market led	Industry-driven	Focused	
Indonesia	Regulatory led	Industry-driven	Focused	
Philippines	Regulatory/Market led	Industry-driven	Focused	
Australia	Regulatory led	Prescriptive	Focused	
New Zealand	Regulatory led	Industry-driven	Wide	

04

The
value
of data

In the Opening Remarks at the Press Conference presenting the EU Open Data Strategy in 2011, Vice-President of the European Commission responsible for the Digital Agenda, Neelie Kroes, gave a speech titled "Data is the new gold".

In the Opening Remarks at the Press Conference presenting the EU Open Data Strategy in 2011[12], Vice-President of the European Commission responsible for the Digital Agenda, Neelie Kroes, gave a speech titled "Data is the new gold". There are countless quotes describing data as the "new gold" or the "new oil", while others state that this is inaccurate since it is the processing of data that creates value from it.

Despite this, the fact that data has an economic value is undisputed and companies that have built their businesses based on their control over vast amounts of data, such as Google and Facebook, enjoy significant premiums in their share values. However, data by itself has little or no value.

An article in CEO Today magazine states that "like gold, data is a commodity. And it is the insight mined from this valuable resource that is the currency that gives you and your company the means to drive organic growth. Therefore, while it is critically important to collect data, the real impact comes from analysing the data you collect and hold in your vault".[13]

---

[12] https://ec.europa.eu/commission/presscorner/detail/en/SPEECH_11_872

[13] https://www.ceotodaymagazine.com/2018/04/is-data-the-new-gold/

Ultimately, the value lies in the metadata and not in the data itself.
Metadata is generated through the analysis of data underpinning a
specific application or use case, and that is ultimately what creates
value to providers and customers alike. The tables in Fig. 7 provide
a comprehensive but not exhaustive list of applications that financial
institutions have been developing from the analysis of data assets.

The majority of the applications listed here fall under Open Banking
and Open Finance wherever the data is accessed through APIs
and under a licensed and permissioned operational framework.
Essentially, it is all about the value to financial services providers
and end users alike. Organisations derive good value from their
investments in data. In one example, "Turkish lender Akbank increased
approved personal loan applications through its mobile app to 30%
from 10% within one short month.... based on a deeper segmentation
of customers"[14] while in Charles Schwab "several departments are
already seeing a return on investment".[15]

Retail branch managers use activity monitoring to analyse customer
activity and identify outreach opportunities, resulting in an improved
customer experience. Lloyds Banking Group rolled out a "machine-
learning algorithm that analyses a user's financial accounts and
transactions to discover intelligent patterns to increase customer
experience". These are a few examples of the benefits of leveraging
structured and unstructured data within the financial services industry.

[14] https://www.fintechfutures.com/files/2018/12/Segments-of-one-customer-insights-in-digital-banking.pdf
[15] https://www.tableau.com/solutions/customer/charles-schwab-equips-more-12000-employees-tableau-advan-ce-data-driven-culture

Analytics-based use cases

(including Open Banking and Open Finance applications)

Fig. 7

USE CASE	DESCRIPTION	UNDERLYING DATA	MARKET	WHO IS OFFERING TO END USERS	EXAMPLES
Account aggregation (balances, lines of credit & utilization, transaction summaries)	**Provides consolidated views** of user's finances, incl. balances & outgoings. Same information can be aggregated **to get better insights into customer** financial needs and behaviour	• Current account transaction data • Credit cards transaction data	🛒 🏪	• Banks • Financial advisors	• Revolut • HSBC • Zurich Group Germany • Allianz AG • Sparkasse
Covid-19 business sector insight	**Forecast** the potential **impact on a company** based on sector and account data	• Sector information • Transaction data • Accounting data	🏪	• Banks • Accounting software providers	• RBS
Cash flow forecasting	Provides **accurate cash flow forecast** based on the actual **payment behaviours** of clients and towards suppliers compared to invoice payment terms	• Transaction data • Invoicing data	🏪	• Banks	• HBSC • NatWest
Accounting automation, payroll & account reconciliation automation	**Automates or improves automation** of otherwise manual and error intensive **accounting** activities	• Transaction data • Invoicing data • Payroll data	🏪	• Banks • Payment companies • Accounting solution providers	• Barclays • Lloyds Banking Group • Santander Deutsche Bank • Idea Bank

🛒 Retail   🏪 Business   🏛 Financial Services

USE CASE	DESCRIPTION	UNDERLYING DATA	MARKET	WHO IS OFFERING TO END USERS	EXAMPLES
Behavioural advisor	Analyse **credit score from payment behavioural data** and **provides** financial behaviour **suggestions aimed to the achievement of a better credit score**	• Account data • Utilities & subscriptions • Prepaid account transactions	🛒	• Banks • Consumer credit companies • Credit bureaus	• RBS • Alex
Marketing intelligence and personalization	Customer rewards to **offer personalized promotions and discounts**. Content is shown in personalized manner through the use of and "AI empowered reccomendation engine"	• Transaction data • Purchcase data • Social media data	🛒	• Banks • Insurance • Retails	• Lloyds Banking Group • Nationwide Building Society • Mercer
Product comparison, information aggregation and recommendation	Aggregate production **information for comparison purpose and provides recommentations based** on questions or analysis of **user behaviour**	• Product data (e.g. account, catd, utility, savings, etc.) • Transaction data • Utility account data	🛒 🏪	• Banks • Insurance • Marketplaces • Processor	• Deutsche Bank
Financial management recommentations	Provision of **financial behaviour advice** based on incoming and outgoing transactions	• Transaction data	🛒	• Banks • Consumer credit companies	• RBS • Sberbank • Tandem Bank • Metro Bank • Santander
Product personalization	Financial behaviour analysis, giving **insights to create new business models** and enable **personalised services and create solutions** for end-customers.	• Transaction data • Behavioural data	🛒	• Banks • Consumer credit companies	• Monzo • Revolut • Sterling • N26 • Akbank

🛒 Retail   🏪 Business   🏛 Financial Services

2/4

USE CASE	DESCRIPTION	UNDERLYING DATA	MARKET	WHO IS OFFERING TO END USERS	EXAMPLES
Reward management optimisation	Optimised **rewards offering** and **management** through **customer profiling** based on transaction and behaviour profiling	• Transaction data • Purchase data • Social media data	🛒 🏪	• Banks • Insurance • Marketplaces • Processors	• Arkea • NatWest/RBS • BNP Paribas • Credit Agricole
Customer journey optimization	**Personalised information mapping** and presentment to **optimise customer online interactions** to a specific outcome	• Transaction data • Behavioural data	🛒	• Banks • Retails	• JP Morgan Chase • BNP Paribas • Charles Schwab • Santander • ABN AMRO
Predictive profiling improvement	Development of more **accurate and effective predictive capabilities** compared to the consented user base alone to provide optimal decision-making	• Anonymised transaction & purchase data • Social media data from consenting & non-consenting consumers • Behavioural data	🛒	• Banks • Retails	• Mastercard
Credit scoring	**Credit score** and **affordability estimate** based on transaction behaviours	• Current account transaction data • Digital footprint • Social media data • Behaviour and psychometric assessment	🛒 🏪	• Banks	• Mynt • Oriental Commercial Joint Stock Bank

🛒 Retail   🏪 Business   🏛 Financial Services

3/4

CRIF DIGITAL   The golden age of data in financial services

USE CASE	DESCRIPTION	UNDERLYING DATA	MARKET	WHO IS OFFERING TO END USERS	EXAMPLES
Pre-delinquency management	**Risk assessment** and **alerts** about accounts at risk of going into arrears and get into debt problems	• Current account data • Credit card data • Savings account data	🛒	• Banks • Consumer credit companies • Credit bureaus	• Bank of Ireland
Fraud prevention/ false positives management	**Complement** the transaction **fraud prevention engine**	• Transaction data • Behavioural data • Geo location	🏪	• Banks • Payment companies • Payment processors	• JustEat • Flixbus • CITI • Goldman Sachs • American Express
Friendly fraud prevention	**Complement dispute management** to prevent friendly fraud	• Transaction data • Behavioural data • Geo location • Social media	🛒 🏪	• Banks • Insurance • Marketplaces • Processor	• Barclays • CHASE • Opt Bank • Sun Finance • City Cash • Cryptocoin • OpenPayd
Anti-money laundering	**Detection** of money laundering activities and **fast adaption to new threats** and new money laundering methodologies. Particularity useful for reducing number of false positives	• Transaction data • Customer data	🏛	• Banks • Insurance	• Danske Bank • United Overseas Bank • HSBC • Commerzbank
Sanction screening	Detects, prevents and disrupts financial crime and sanction risk by **reducing the number of false alerts** and **staying up to date to regulatory changes**	• Transaction data • Customer data • Sanction entity lists	🏛	• Banks • Insurance	• Danske Bank • Intesa Sanpaolo

🛒 Retail   🏪 Business   🏛 Financial Services

4/4

05

Three cases from Crif Digital

01 Top tier European consumer credit organisation: full digital data and process integration

A top tier European consumer credit organisation doubled the amount of approved loans for the same number of website visits and defaults. This was achieved through an 88% reduction in declined applications compared to their pre-PSD performance and a 68% decrease compared to their initial PSD2 implementation.

The overall application has seen a 50% increase in the number of automated loan decisions compared to the previous amount of manual reviews and overrides, while improving the customer experience by halving the time taken by consumers to fill in the application form.

The project was developed to improve risk assessment and management with a particular focus on mid-tier risks, while automating the process with a full digital approach. The solutions included a tighter integration of scoring from current account data within a wider credit score and the integration of upgrade and downgrade rules to lower the overall risk within this tier. Fig. 8 illustrates the process with the areas in which data feeds and controls have been improved.

PSD2 integrated into full digital process with a European consumer credit company

Fig. 8

LOAN CHOICE
Amount, installment, purpose

DATA COLLECTION
Subject data and credit line data

PSD2 POP-UP
This shows the benefits of sharing current account information

ADDITIONAL DATA
Request for payslip and IBAN

ADDITIONAL DATA
Automated computation of salary and IBAN via PSD2

TRADITIONAL FLOW
From compilation:
~ 15 minutes

PSD2 FLOW
From compilation:
~ 8 minutes

CONTROLS
AF/AML, traditional scoring

CONTROLS
AF/AML, traditional scoring + PSD2 data

ASSESSMENT

DIGITAL SIGNATURE
If approved

02 European credit card issuer: financial inclusion and increased approval

A European credit card issuer wanted to increase the approval of applications and to play a role in supporting financial inclusion in the countries it operates in. This was achieved by reaching out to applicants that would otherwise have been declined. Nearly half of those invited agreed to share access to their current account data with their application being reassessed through the process illustrated in Fig. 9.

This project, now with the configuration fully operational after a successful pilot, highlighted a large share of creditworthy applications that would otherwise be declined, in particular within the "new2credit" and "new2country" demographic segments. 27% of those that were at first potentially declined have now been approved, with a sizable increase in the number of cards issued without any negative impact on the risk of outstanding balances.

Monoline credit card issuer – application re-scoring process

Fig. 9

BPO contacts the end user by **phone** to enable access to accounts for reassessment

The **BPO** operator automatically sends an e-mail to the end user containing the **XS2A link**

The **BPO** operators receives a **notification** when the PSD2 report is **ready** to download

The **application** is reassessed using a scorecard that considers the **PSD2 SCORE** tranches and KPIs

CRIF AISP navigation widget for XS2A

03 Top tier European retail and SME bank: lending campaign for existing customer base

This top tier European retail and SME bank wanted to increase cross-selling to its existing customer base by leveraging the current account information held on its books. Key to this use case is NEOS CATCH, CRIF Digital's transaction categorisation engine, which is used to develop a creditworthiness assessment based on balance bands, income bands, monthly available income and on a score, which is centred around current account transaction information (NEOS Score). This has been leveraged through the application of KPIs aimed at selecting customer profiles consistent with cross-selling propositions - all deployed through several ongoing, parallel, ad-hoc campaigns. Fig. 10 illustrates the selection process of customers for proactive, pre-approved upselling.

The AI-based categorisation of current account transactions and the marketing profiling aimed at identifying higher offer redemption priority led to sound results, with 49% of the customer base responding positively to at least one offer. This has led to a 3.0x redemption on personal loans, 3.3x on deposits and savings accounts, and 2.8x on bank assurance offers.

The AI-based analysis of account transactions also highlighted secondary customer accounts that, despite receiving salaries, were not used as the primary account. This led to a further re-evaluation of the current account portfolio and a campaign to win back customer activities on the account.

Top tier European retail and SME bank – proactive, pre-approved upselling

01 The bank wants to find a way of identifyng the most reliable customers to propose credit products that are appropriate to their current needs

02 Using NEOS Inisight, the bank analyzes the entire portfolio on a monthly basis and obtains updated scores and KPIs for each customer

03 The bank selects the list of eligible customers to be pre-approved. Current account KPIs are further drivers for the calculation of the credit limit

04 The bank can quickly identify the most reliable and responsive customers

05 The bank notifies the customer fort the availabiliy of a new loan through its touchpoints (e.g. PFM/BFM)

Transforming unrefined data into "gold"

Structured data is easily searchable, but this does not exploit its full potential. Data needs refining and processing using sophisticated technologies.

For example, to fully leverage and make sense out of unstructured data becoming accessible through Open Banking, it is fundamental to use advanced analytics applications such as AI-based categorisation engines, including machine learning for gradient boosting and deep neural networks.

These engines can apply AI text mining to detect and classify data e.g. for profiling, customer segmentation or other commercial purposes. However, the combined ability to manage very large data volumes with the availability of sophisticated AI technologies is not something commonly available in the current market. The early development stage of the Open Banking market is limiting these capabilities to only a few providers and, even amongst those, there is still a significant capability gap and, in some cases, significant unexploited processing potential.

CRIF Digital's multilingual categorisation engine, based on CRIF's taxonomy with machine learning and AI detection methodologies, enables a very detailed and accurate identification of spending categories and classifies individual account transactions, as indicated in Fig. 11.

[23] Software as a Service

CRIF advanced categorisation engine expertise

Fig. 11

SHARPENING THE LEARNING CURVE
Through Augmented Reality technology

PERFORMANCE
Accuracy > 91% with over 10,000 identified terms

THRESHOLD
Ability to set custom thresholds

RULES BASED CATEGORIZATION
Ability to use expert rules upfront, upftream to the Machine Learning algorhytm

SCALE AND DIFFERENTIATION
Over 240 identified categories based on the transaction type (including inbound and outbound) and the type of customer (e.g. retail / business)

FOCUS AND SPECIALIZATION
The categories are designed so to identify suitable behaviours needed to optimise credit decisioning and marketing opportunities

MULTILANGUAGE
The engine is designed to operate across multiple languages making it able to be deployed in multiple and cross country

BANK CUSTOM DESCRIPTION TRANSLATION
The categozation engine interprets the different descriptors in use by different organisations across transactions of similar nature

There are many possible applications of the categorisation engine.
The most intuitive and straight-forward is to enrich current account
holders' information, which is used as a baseline for a number of
products. These include, for example, lending, profiling and, more in
general, customer management.

**The "categorised" current account can offer insights to micro-
segment on an individual customer level**, something that would
otherwise be lost when only looking at customers through the lens
of segmentation. For example, based on a CRIF Digital project with a
European bank, by analysing business segment expenditure, it was
discovered that a small percentage of this referred to "training".

**By investigating this insight, CRIF Digital showed that companies
with this expenditure type were going through a potential M&A
initiative** and, as such, in need of specific services (financial,
consulting, insurance, etc.) from their banking provider.
See Fig. 12 for an illustration of one of the outcomes of the
analysis using this application.

Illustration from CRIF Digital's categorisation engine application, highlighting an insight into transaction behaviour that would otherwise have gone unnoticed

Fig. 12

Amount (negative) has been converted to its absolute value

☐ 43,1% Other Expenses   ■ 7,9% Wages & Salaries   ■ 2,7% VAT

■ 2,4% Maintenance   ■ 2,3% Taxes   ■ 3,3% Catering & Restaurant

■ 2,1% Training   ■ 2,1% Raw Materials, Goods &   ■ 2,3% Mortgage and Loan OUT
                                   Service
                                                                   ■ 2,0% Professionals

Illustration from CRIF's categorisation engine application to the corporate portfolio, providing insights into the expenditure mix within a target industry
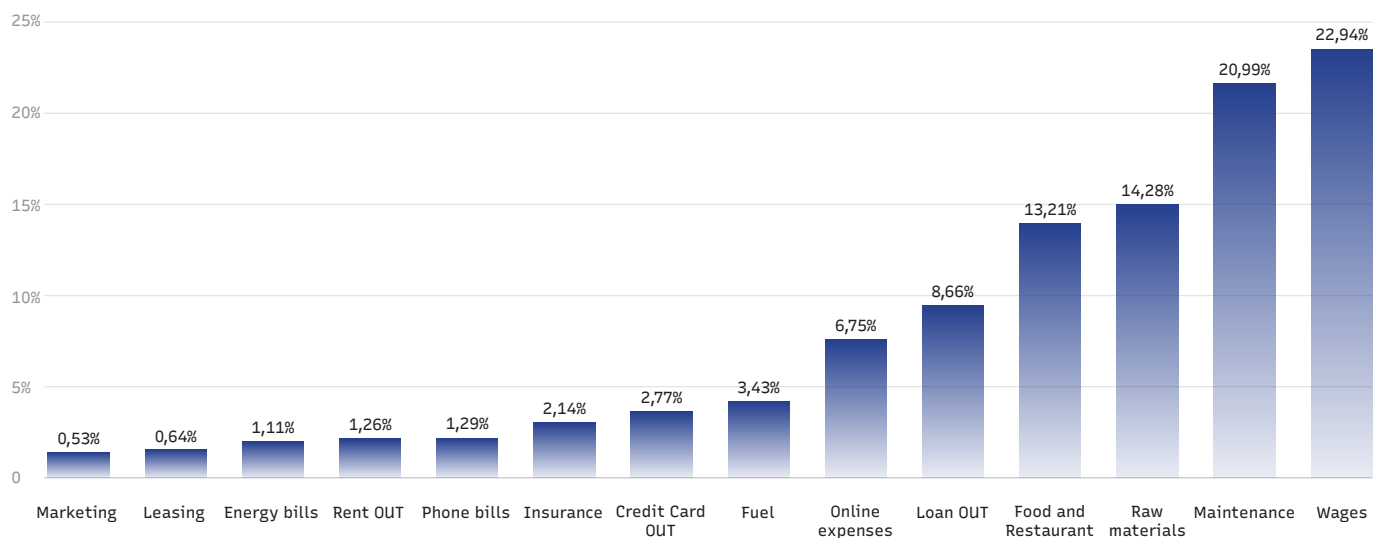
Fig. 13

Based on our experience, it is the less intuitive and indirect insights that are often the most valuable. The information that is now accessible through Open Banking can feed valuable customer insights in real time, using the categorisation engine.

For example, through an aggregated analysis of expenditure transactions, it is possible to gain insights into what is typical for each industry. Fig. 13 provides an illustration of the analysis applied to the expenditure mix within the construction industry.

This provides insights that can be used to define and fine-tune value propositions or sales initiatives aimed at companies within this industry. Otherwise, by comparing how this mix evolves over time, it is possible to gain insights into how this industry is evolving.

Industry: Construction

Categorization of the most important expenses

	Marketing	Leasing	Energy bills	Rent OUT	Phone bills	Insurance	Credit Card OUT	Fuel	Online expenses	Loan OUT	Food and Restaurant	Raw materials	Maintenance	Wages
%	0,53%	0,64%	1,11%	1,26%	1,29%	2,14%	2,77%	3,43%	6,75%	8,66%	13,21%	14,28%	20,99%	22,94%

Data can provide highly valuable insights. However, different technologies need to be applied while also keeping in mind compliance considerations. For example, since self-learning technologies are black boxes, the decision-making process is untransparent. Although widely used, the application of AI solutions to credit scoring using unstructured data is currently a grey area. Many machine and deep learning techniques fail to comply with financial services regulations, especially in the context of creditworthiness assessments.[16]

Regulatory bodies have highlighted the need for more research to improve the interpretability of AI algorithms to mitigate risks such as bias and discrimination.[17-18] "Understanding of the input data is important to strengthen trust and to ensure that personal data are lawfully used and the risk of amplifying bias in historical data is addressed."[19] Many research studies[20] consider the interpretation of algorithms to help understand the decision process by studying so-called model-agnostic techniques such as local interpretable model-agnostic explanations (LIME). Leveraging its network of data scientists and as a close partner of universities, CRIF is driving the analysis of such technologies to fulfil regulations and to test methodologies allowing unambiguous explanations of machine learning models.[21-22]

[16] GDPR, Art. 22 (3) https://www.privacy-regulation.eu/en/article-22-automated-individual-decision-making-inclu-ding-profiling-GDPR.htm and Law Enforcement Directive, Art. 11 (1) https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016L0680&rid=1

[17] European Parliament (2019), EU Guidelines on ethics in artificial intelligence: Context and implementation, https://www.europarl.europa.eu/RegData/etudes/BRIE/2019/640163/EPRS_BRI(2019)640163_EN.pdf

[18] European Commission. 2018a. "Communication Artificial Intelligence for Europe", file:///C:/Users/ClaudeFabienne-D%C3%BCrig/Downloads/CommunicationArtificialIntelligence.pdf

[19] The World Bank Group (2019), Credit Scoring Approaches Guidelines, http://pubdocs.worldbank.org/en/935891585869698451/CREDIT-SCORING-APPROACHES-GUIDELINES-FINAL-WEB.pdf

[20] Financial Stability Board (2017), Artificial intelligence and machine learning in financial services https://www.fsb.org/2017/11/artificial-intelligence-and-machine-learning-in-financial-service/

[21] Giorgio Visani and Federico Chesani and Enrico Bagli and Davide Capuzzo and Alessandro Poluzzi (2020) Explana-tions of Machine Learning predictions: a mandatory step for its application to Operational Processes https://arxiv.org/pdf/2012.15103.pdf

[22] Giorgio Visani and Enrico Bagli and Federico Chesani and Alessandro Poluzzi and Davide Capuzzo (2020) Statistical stability indices for LIME: obtaining reliable explanations for Machine Learning models https://arxiv.org/pdf/2001.11757.pdf

Economic context and short - term trends

Covid-19 is affecting everybody's life around the globe. Despite the increasing challenges for individuals, the pandemic also poses obstacles for companies.

High liquidity requirements, the elimination of traditional distribution channels, and a whole new working environment for employees are just a few of the challenges that most companies are confronted with. All this while governments are keen to protect their industrial bank, related jobs, and economic output. Banks have become a channel for government subsidies and have to face the challenge of managing their own business and supporting their customers in difficult times.
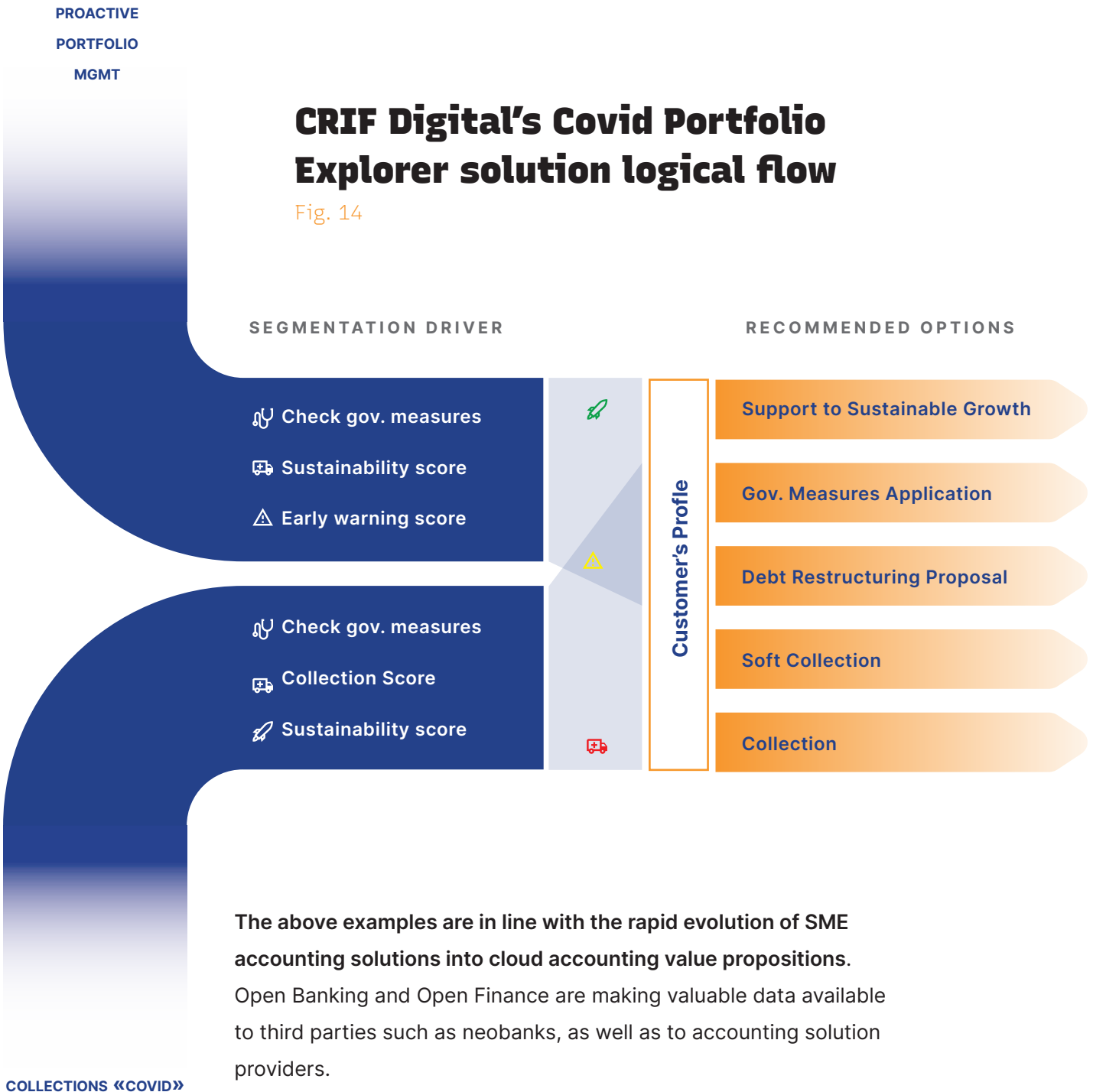
While all banks globally focus their SME and corporate banking considerations on their customers' cash flow positions, some leverage data and analytics with dedicated solutions. For example, the Royal Bank of Scotland, in cooperation with accounting software provider FreeAgent, offers customers the opportunity to forecast critical business cash flow scenarios and present potential outcomes of cost adjustments to business.

HSBC has a similar approach and supports its customers with a cash flow tool, which can generate forecast horizons and test cash flow scenarios. In Germany, Penta, in cooperation with Banxware, an SaaS[23] service provider, is offering the first fully digital online application for government support. A major European bank and the consumer credit branch of a major French banking group have also rolled our similar initiatives with the support of CRIF Digital analytics.

Our value propositions range from industry sector risk exposure scores, "Covid Portfolio Explorer" and "Pulse", to development solutions for companies to improve their financial management down

[23] Software-as-a-Service

to supporting lenders in collection, taking into account the nature, timing and size of government support initiatives across the European markets. Fig. 14 provides an illustration of the logical flow of the "Covid Portfolio Explorer" solution.

PROACTIVE PORTFOLIO MGMT

CRIF Digital's Covid Portfolio Explorer solution logical flow

Fig. 14

SEGMENTATION DRIVER

RECOMMENDED OPTIONS

- Check gov. measures
- Sustainability score
- Early warning score

- Check gov. measures
- Collection Score
- Sustainability score

Customer's Profile

Support to Sustainable Growth

Gov. Measures Application

Debt Restructuring Proposal

Soft Collection

Collection

COLLECTIONS «COVID»

The above examples are in line with the rapid evolution of SME accounting solutions into cloud accounting value propositions. Open Banking and Open Finance are making valuable data available to third parties such as neobanks, as well as to accounting solution providers.
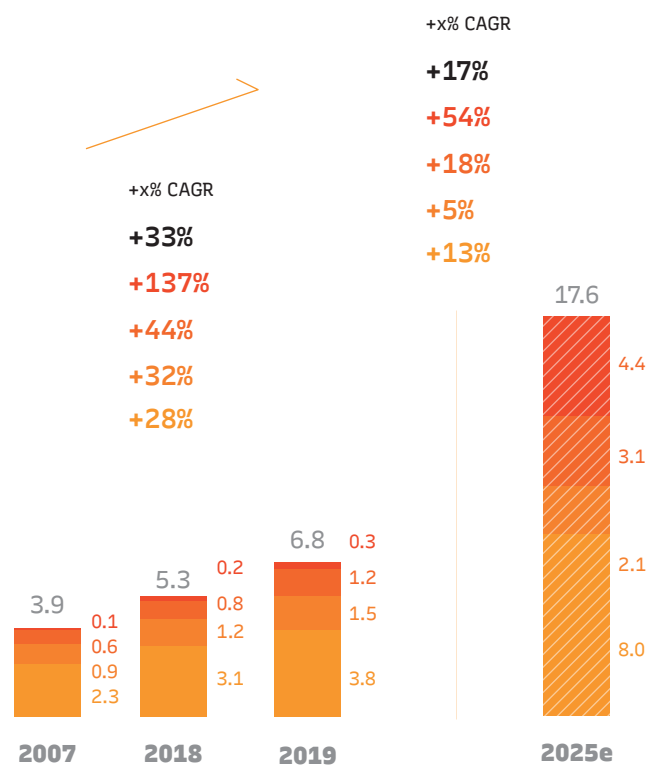
The latter have seen an increase in demand and worldwide adoption in terms of subscriptions and the number of SMEs moving to cloud-based solutions, in particular with smaller companies that were previously opting for specialist professional support. Fig. 15 and Fig. 16 provide a view of the growth dynamics in the cloud accounting industry.[24]
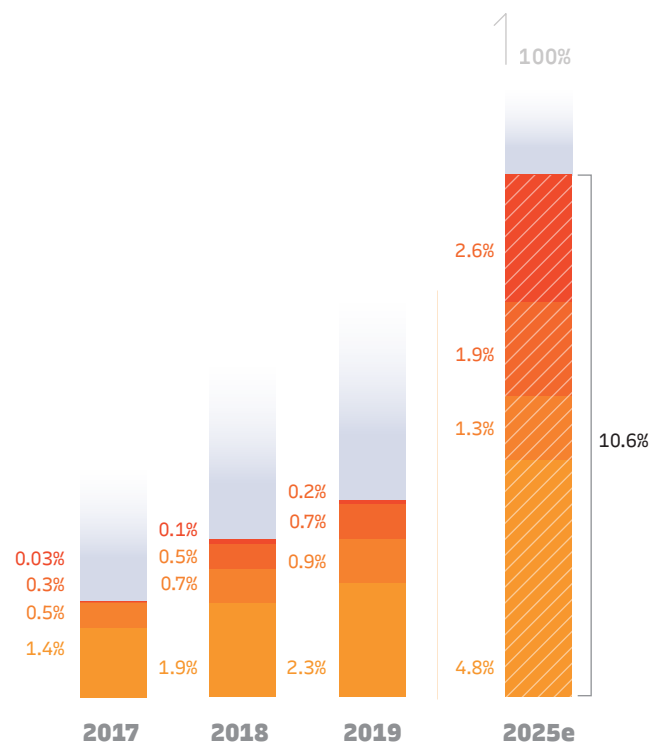
Trends in the cloud accounting industry

Fig. 15 and Fig. 16

Number of subscriptions (in million pcs)

+x% CAGR
+17%
+54%
+18%
+5%
+13%

+x% CAGR
+33%
+137%
+44%
+32%
+28%

	2007	2018	2019	2025e
Total	3.9	5.3	6.8	17.6
Row	0.1	0.2	0.3	4.4
UK	0.6	0.8	1.2	3.1
Australia & New Zealand	0.9	1.2	1.5	2.1
North America	2.3	3.1	3.8	8.0

Number of SMEs using Cloud Accounting vs. total number of SMEs (%)

100%

	2017	2018	2019	2025e
Row	0.03%	0.1%	0.2%	2.6%
UK	0.3%	0.5%	0.7%	1.9%
Australia & New Zealand	0.5%	0.7%	0.9%	1.3%
North America	1.4%	1.9%	2.3%	4.8%

10.6%

☐ Total WorldWide [(1)]   ■ Row [(2)]   ■ UK   ■ Australia & New Zealand   ■ North America

[24] Note: Adoption rates are based on the number of subscriptions to the two dominant cloud accounting players for SMEs, Intuit and Xero. 100% refers to the total number of SMEs in each region

(1) Includes numbers for the EU, China, Brazil, India, Russia, Australia, New Zealand, the UK and North America;
(2) Rest of the World includes Brazil, India, China, Russia and the EU Source:

Xero annual reports and company webpage; Intuit annual reports and company webpage; New Zealand Ministry of Business, Innovation & Employment; National Federation of Self Employed & Small

Businesses Limited; Office of the United States Trade Representative; International Labour Organization; World Bank; Eurostat; Accounting web; Accaglobal; Accountancy daily; CRIF analysis

These solutions are highly complementary to any cash flow service offered by a bank to its SME customers and, thanks to the potential of unstructured data, offer the ability to develop into value-added services, such as SME cash management, reconciliation automation, and payroll services.

While bringing added value to customers, this generates business opportunities for the financial services industry for core business and additional added-value service expansion.

08

Conclusions - releasing the value of data

As new technologies enable full access to unstructured data, new opportunities for financial services are emerging, especially in the era of Open Banking and Open Finance.

Opportunities for banks and insurers are almost unlimited and the use of data will be a strong differentiator in the future competitiveness of financial institutions. The experience gained through CRIF Digital with clients achieving success through an increased redemption rate, better risk management and risk profile quality of their lending portfolio, to name just a few, is proof of the potential value held in unstructured data and in the capabilities of data analytics and artificial intelligence.

Drawing on our 32 years' experience in data insights and analytics, there are a few valuable lessons learned that should be considered when looking at deriving value from data:

01   **The importance of the availability, quality and relevance of data** – without heterogeneous and sufficient data to start from, it is not possible to speed up the development of all the analytics necessary to leverage this new information "gold mine".

From the experience gained through CRIF Digital, a key step is to develop an efficient categorisation engine and any other type of analytics algorithm. Data analytics models cannot differentiate between causation and correlation. Hence, poor data quality will skew results when performing unstructured data analysis. Pre-processing data is crucial to making unstructured data easier for machines to analyse and to increasing its quality.

02 **The value of data increases as the data available for processing increases** - Open Banking is currently the most talked about opportunity for deriving value from data. While this is based on valuable product-related data pools, there are other complementary unstructured and dark data pools to be leveraged.

From CRIF's experience, data sources rarely substitute each other, but in almost all cases they complement each other, greatly improving the value of the analytics. This has a quantitative and qualitative explanation.In general, on the quantitative side, we have observed a low correlation (below 0.5) between structured and unstructured data.

The independence (technically referred as "orthogonality") of two data sources is a prerequisite for generating, if combined through a scorecard, a quality improvement in profiling and forecasting. CRIF Digital has shown this through the combination of Open Banking data with its credit bureau information.

On the qualitative side, it is the unstructured data that makes it possible to enrich the "traditional" data source with new information. In early warning processes, it is only after the introduction of the text mining capabilities of a categorisation engine that it is now possible to get a better understanding of customer behaviour.

For example, by spotting if the account holder is going into overdraft as a result of losing his/her job (e.g. no more salary transactions) or because he/she has simply started to spend more and what he/she is spending it on. This is a very important factor that suggests combining structured and unstructured data also to guarantee the best analytics independently of the data assets available in different geographies.

03   **Ultimately it is all about value to the end user** – consumers and businesses expect a seamless digital experience. It is important to zero in on what matters most to end users: speed and flexibility, transparency, trust, interaction and engagement.

Data-driven solutions should be embedded within existing workflows so that they are easily adopted as process enhancements. Similarly, end users do not need to recognise the term 'Open Banking' to understand its benefits. Ultimately, it is all about value to users and how the analysis of customer data can improve their financial lives.

09

About
Crif Digital

CRIF Digital – a brand of CRIF Group – is a platform with solutions exploiting the powerful opportunities generated by Open Banking.

CRIF Digital solutions speed up the business digital transformation, better performances and offer a smooth user experience to final customers.

Ensuring a memorable digital journey for customers.

We help customers who want to consciously manage their finances and have a smart, safe and personalized relationship with the financial institution, so that they can **save time and effortlessly find an answer to all their needs**, even from a distance. The solutions we provide make them feel safe when online, receive truly personalized offers and perform tasks in a pleasant and easy way.

Leveraging customer data-driven approach to increase long-term value for Financial Institutions.

We help Financial Institutions to build long-lasting and meaningful relationships with their clients, designing digital experiences and services. Customer experience is an important aspect to consider, but it is part of a more integrated strategy based on digital solutions we can provide to generate value for businesses.

Increase the client value unlocking a better and deeper knowledge of the customer base to establish a new paradigm of relationship: not pushing products but foreseeing and matching the customer's needs.

For more info:

https://www.crif.digital

crif.digital@crif.com