The use of Non-Traditional Data in Financial Services
Agenda

Non-traditional data landscape

Uses of Non-traditional data by Fintechs

South African Fintech survey insights

Considerations and implications for Financial Regulators
Big data has become vast and ubiquitous

Big data is growing rapidly...

- **7000 Petabytes** of data stored across the globe
- **5 billion** mobile phones currently in use
- **30 billion** pieces of content shared on Facebook every month
- **40%** projected growth in global data generated per year
- **20 hours** of video content uploaded per second
- **2.9 million** Emails sent every second

...with Financial Services being one of the leaders

Ease of Capture

- **High**
  - Utilities
  - Healthcare
  - Financial Services
  - Telecoms, Media & Tech
- **Low**
  - Manufacturing
  - Consumer Electronics
  - Travel, Transport, & Logistics
  - Professional Services
  - Accommodation & Food
  - Natural Resources
  - Real Estate
  - Retail
  - Administrative Services
  - Educational Services
  - Construction
  - Government
  - Educational Services
  - Construction
  - Government

Value

- **Low**
  - Financial Services
  - Telecoms, Media & Tech
- **High**
  - Utilities
  - Healthcare
  - Manufacturing
  - Consumer Electronics
  - Travel, Transport, & Logistics

Source: Dataconomy.com; McKinsey & Company
Organisations are increasingly leveraging non-traditional data sets to capture value

Non traditional data refers to data sourced from external sources used to supplement core internal organisational data.

Traditional data refers to data gathered from internal sources/core systems.

Fintech’s are making use of external data like social site postings, search engine keywords, online transactions, photo albums, and group chats to gather insights about consumers’ preferences and provide cross-sell/targeted ads.

Telecom companies are analyzing internal unstructured data like call records, and URL/contents to reduce churn (e.g., call centers’ reps’ behavior, key phrases that led to churn like “sorry, I can’t help you with that”).

Non-traditional data refers to data sourced from external sources used to supplement core internal organisational data.

Traditional data refers to data gathered from internal sources/core systems.

Structured data type

Unstructured data type

Credit scores
Property ownership
Call patterns
Location check-in data
Professional relationships
Behavioral characteristics
Marketing trends
Medical trends
Medical trends
Personal relationships

Non-Traditional

Structured

Financial Holdings
Customer profile
Transactions
Email derived relationship data
Location data (on bank properties)
Channel behaviors (on bank properties)
Employee sentiment data
Call records
Channel usage
Purchasing patterns

Traditional

Unstructured

Home equity
Musical tastes
Items purchased
Viewing behaviors
Items purchased
Market trends
Medical trends

Fintech’s are making use of external data like social site postings, search engine keywords, online transactions, photo albums, and group chats to gather insights about consumers’ preferences and provide cross-sell/targeted ads.

Telecom companies are analyzing internal unstructured data like call records, and URL/contents to reduce churn (e.g., call centers’ reps’ behavior, key phrases that led to churn like “sorry, I can’t help you with that”).

Non-traditional data refers to data sourced from external sources used to supplement core internal organisational data.

Traditional data refers to data gathered from internal sources/core systems.

Structured data type

Unstructured data type

Credit scores
Property ownership
Call patterns
Location check-in data
Professional relationships
Behavioral characteristics
Marketing trends
Medical trends
Medical trends
Personal relationships

Non-Traditional

Structured

Financial Holdings
Customer profile
Transactions
Email derived relationship data
Location data (on bank properties)
Channel behaviors (on bank properties)
Employee sentiment data
Call records
Channel usage
Purchasing patterns

Traditional

Unstructured

Home equity
Musical tastes
Items purchased
Viewing behaviors
Items purchased
Market trends
Medical trends

Data type
There are several stages involved in converting non-traditional data into insight:

<table>
<thead>
<tr>
<th>Stage</th>
<th>Description</th>
<th>Technologies</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Data Collection</td>
<td>Gathering of quantitative and qualitative information on specific variables with the aim of evaluating outcomes or gleaning actionable insights.</td>
<td>API: direct database-to-database data transmission enabling granular, real-time reporting and automated validation.</td>
</tr>
<tr>
<td>2. Data Cleaning/processing</td>
<td>Preparing of data for analysis by removing or modifying data that is incorrect, incomplete, irrelevant, duplicated, or improperly formatted.</td>
<td>Machine Learning: enables automated data analysis, anomaly detection, merge-sort, scoring and other functions.</td>
</tr>
<tr>
<td>3. Data Storage</td>
<td>The recording (storing) of information (data) in a storage medium.</td>
<td>Cloud computing: enables the storage of huge volume of data.</td>
</tr>
<tr>
<td>4. Data Analytics</td>
<td>Data analytics is the science of analyzing raw data in order to make conclusions about that information.</td>
<td>Hadoop: It provides a software framework for distributed storage and processing of big data.</td>
</tr>
<tr>
<td>5. Insight/Use-Case</td>
<td>FSPs are able to convert non-traditional data into insight to derive benefits such as financial inclusion, personalization, customer experience and affordability.</td>
<td>Machine Learning: makes it possible to quickly and automatically produce models that can analyze bigger, more complex data and deliver faster, more accurate results.</td>
</tr>
</tbody>
</table>

FSPs are able to convert non-traditional data into insight to derive benefits such as financial inclusion, personalization, customer experience, and affordability.
There are a number of technologies and techniques used to derive value from Traditional & Non-traditional data

1. Data
   - Identify, combine, and manage multiple sources of data
   - Creatively source internal and external data
   - Upgrade IT architecture and infrastructure for easy merging of data

2. Technologies and techniques used to process Traditional data & Non-Traditional data
   - **Data lakes**: scalable storage solution for diverse structured, semistructured, and unstructured data
   - **Web portal**: static file upload via web site with built-in automated validation checks
   - **Chatbots**: automated capture and interpretation of qualitative data enabling data collection in real time
   - **Application programming interface (API)**: direct database-to-database data transmission enabling granular, real-time reporting and automated validation
   - **Data cubes**: granular data storage and transmission solution enabling real-time data collection
   - **Web scraper**: automated capture of web data by “bots”
   - **Cloud computing**: on-demand network access to share computing resources (e.g., networks, servers, storage, applications, and services)
   - **Distributed ledger technology (DLT)**: a network to securely propose, validate and record changes to a synchronised ledger distributed across multiple nodes
   - **Robotic process automation (RPA)**: partial or full automation of manual, rule-based and repetitive human activities by “bots”
   - **Dashboards**: customisable, dynamic interactive reporting tools that automatically fetch and render data in meaningful and actionable visualisations
   - **Text mining**: automated extraction of meaning from textual data
   - **Machine learning**: automated data analysis enabling anomaly detection, merge-sort, scoring and other use cases
   - **Geographic information systems (GIS)**: automated analysis of spatial or geographic data

3. Execution
   - Delivering analytical insight by creating simple, understand-able tools/campaigns for people on the front lines
   - Update processes and develop capabilities to enable tool use
Our Case studies focused on Payments, Lending, and Insurtech, where non-traditional data is most prominent

<table>
<thead>
<tr>
<th>Key segments using non-traditional data</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Payments</strong></td>
</tr>
<tr>
<td>Payments fintechs are using non traditional data with their extensive transaction data to enhance mobile payments user experience</td>
</tr>
<tr>
<td><strong>Lending</strong></td>
</tr>
<tr>
<td>Lending Fintechs are using non-traditional to build more accurate scorecards, improve customer profiles, make better credit decisions, and to manage overall credit risk</td>
</tr>
<tr>
<td><strong>Insurtech</strong></td>
</tr>
<tr>
<td>Insurtechs are using non-traditional data to offer policyholders better premiums and to manage better risk</td>
</tr>
</tbody>
</table>
Agenda

- Non-traditional data landscape
- **Uses of Non-traditional data by Fintechs**
- South African Fintech survey insights
- Considerations and implications for Financial Regulators
In Payments, non-traditional data is being used to gain greater understanding of customer needs and develop products.

**Use-cases**

- **Visa’s joint venture with the Gap** is an early example of using real-time, location-based Payments data to make targeted customer offers.
- **Mastercard partnered with Banks** in using their Payments data to formulate a 360° view of the customer to drive insights unavailable through isolated data sources.
- **Yoco leverages Payments POS data** generated through their devices to offer SMEs working capital loans.

**Benefits**

- **Mobile payments user experience**, technical quality, and acceptance network improved significantly.
- Actionable insights leading to **increased personalization and improved value propositions**.

**Risks**

- **Privacy and data protection** concerns. This stems primarily from the fact that much of non-traditional data is often linked with personal information.
In Lending, non-traditional data is being used to enhance the credit lifecycle

**Use-cases**

- **Lenddo** aggressively using advanced machine learning to comb through vast sources of alternative data to predict an individual's creditworthiness
- **Credit sesame** uses behavioral data points to grade creditworthiness and offer consumer loans
- **Lendix** uses non-traditional SME data and doesn't require financial statement to grant loans to SMEs
- **Rainfin** collects data from a broad basket of databases to assess creditworthiness of their clients

**Benefits**

- Drives financial inclusion for creditworthy consumers previously excluded
- Promotion of access to finance for small and medium enterprises (SMEs)
- Helping underserved groups by enhancing financial networks

**Risks**

- Potential for bias on the basis of sensitive demographic characteristics
- The process makes it difficult for consumers to verify or challenge unfair decisions
- Alternative credit assessment not always reliable
- Data privacy and data protection challenges
Insurtechs are using non-traditional data to offer products and services tailored for consumer needs

Use-cases

- **Discovery Insure and Root Insurance** use connected IoT-based devices and telematics to encourage positive driving behavior and obtain more data for better premium pricing.
- **Flyreel** provides AI-powered solutions that offers customers and their insurance carriers visibility into property features. It further uses AI for underwriting of commercial and residential properties.
- **Cuvva** uses connected IoT devices to offer consumers insurance that adapts to their behavior and can also be customized and switched on an off by consumers on demand.

Benefits

- **Personalisation** – Non traditional data enables Insurtechs to offer personalized customer journeys and customizable product options.
- **Incentivisation** - Through incentives Insurtechs using non-traditional data encourage positive behavioral change.

Risks

- **Privacy and data protection concerns** stemming from the misuse and mishandling of non-traditional data without consent.
- **Discrimination** - More granular data can lead to the exclusion of some consumers from being insured.
In summary, our Case Studies revealed that Non-traditional data is being used to drive several key benefits in the financial sector:

<table>
<thead>
<tr>
<th>Fintech Products &amp; Services</th>
<th>Explanation of potential data / analytics impact</th>
<th>Potential scale of impact</th>
<th>Examples</th>
</tr>
</thead>
</table>
| **Affordability**           | • Lending product  
                              | • Insurance product  
                              | • Liquidity product  | • Non-traditional data and advanced analytics can lower delivery costs for many financial products, particularly those that entail some form of risk assessment (e.g., lending, insurance)  
                              |                                                    | • Lower delivery costs will allow Fintechs to lower prices and make products more affordable to low-income consumers  |
| **Awareness & understanding** | • Insurance product  
                              | • Liquidity product  | • Analytical modeling can help to identify groups of consumers that will be most receptive to marketing and education campaigns  
                              |                                                    | • Analytics can help to determine which messages are likely to resonate most with consumers  |
| **Accessibility**           | • Transaction / deposit product  
                              | • Lending product  | • Data on patterns of consumer geo-location and mobility can help companies determine where to locate operations and how best to reach consumers  |
| **Desirability**            | • Liquidity product  | • Analyses of data that suggest consumer behaviors and preferences (e.g., census data, social media data) can help companies develop products that are likely to meet the financial needs of the poor  |
Agenda

Non-traditional data landscape

Uses of Non-traditional data by Fintechs

South African Fintech survey insights

Considerations and implications for Financial Regulators
## Survey Findings

<table>
<thead>
<tr>
<th>Data management</th>
<th>Benefits</th>
<th>Risks</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Non traditional Data (NTD) sources (i.e. GPS, Smartphones, Social Media, IoT)</td>
<td>4. Delivering personalisation through NTD</td>
<td>8. Data protection risk management</td>
</tr>
<tr>
<td>2. Non traditional Data (NTD) pulling methods (i.e. Screen Scraping, Hadoop)</td>
<td>5. Delivering financial inclusion through NTD</td>
<td>9. Data Privacy risk management</td>
</tr>
<tr>
<td>3. Non traditional Data (NTD) processing methods (i.e. AI/ML/RPA)</td>
<td>6. Delivering enhanced customer experience through NTD</td>
<td>10. Decision bias risk management</td>
</tr>
<tr>
<td></td>
<td>7. Delivering lower prices through NTD</td>
<td>11. Data transparency</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>0. Low</th>
<th>1. Basic</th>
<th>2. Good</th>
<th>3. Best-in-class</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Key benefits & risks findings

#### Key Benefits

- **Personalisation**: FSPs are converting non-traditional data into insight to design products and services that meet individual specific requirements/needs

- **Financial inclusion**: FSPs are able to use non-traditional data to drive the uptake, usage and quality of their products and services, resulting in access to financial services to the previously underserved groups

- **Customer experience**: FSPs are able to use non-traditional data to drive the interaction between themselves and customers

- **Affordability**: the use of non-traditional data by FSPs is enabling them to design cheaper and better services to consumers

#### Key Risks

- **Discrimination**: granular data and algorithms may result in consumers being excluded from accessing certain products and services due to real or perceived risks

- **Data privacy**: consumers data being accessed and used without consent

- **Data protection concerns**: lack of safeguards in place to prevent data leaks and misuse by third parties

- **Data misuse**: inappropriate use of consumer data beyond the scope of their consent

- **Transparency**: Customers may not know how FSPs are collecting, using and sharing their data, and what benefits they are getting

- **Fairness**: Consumers have no safeguards against unfair discrimination and exclusion resulting from their data. In some cases customers may have no right to correct or update their data
Agenda

Non-traditional data landscape

Uses of Non-traditional data by Fintechs

South African Fintech survey insights

Considerations and implications for Financial Regulators
## Six Key Considerations (1/2)

<table>
<thead>
<tr>
<th></th>
<th>Disclosure &amp; Informed Consent</th>
<th>Security</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><strong>FSPs should be clear about their use of customer data, attain customer agreement to their customer data policies and, where appropriate, seek consent for specific uses</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• <strong>Informed consent</strong>: FSPs need to provide clear and accessible information about how customer data will be used (e.g. terms and conditions).</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• <strong>Transparency</strong>: Customers should be able to view or know the data that are collected about them, how they are used and whether they are shared with a third party.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• <strong>Ability to revoke consent</strong>: Customers should be able to request that data about them no longer be used by a FSP (e.g. the right to be forgotten).</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• <strong>Legitimate use</strong>: FSPs may not need to seek consent when using data for legitimate interests (e.g. those required by law).</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td><strong>FSPs should be held responsible and accountable for data security</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• <strong>Liability</strong>: A clear liability framework should be in place that ensures the responsible party is held accountable for data security and for harms caused by breaches of its respective data security duties of care.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• <strong>Traceability</strong>: FSPs need to be able to identify where data were improperly used or accessed in the event of a security breach.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td><strong>FSPs should to disclose to customers which of their data points they are using and enable customers to intervene and limit use where applicable</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• <strong>Intervention</strong>: Customers should be able to intervene to gain information or limit use of data they control, and FSPs should respond appropriately.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• <strong>Limited use</strong>: Where reasonable, a maximum time period that data can be retained by FSPs should exist, as well as limits on certain sensitive data types or uses.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Six Key Considerations (2/2)

<table>
<thead>
<tr>
<th></th>
<th>Portability</th>
<th>Privacy &amp; Data Misuse</th>
<th>Algorithms &amp; Analytics</th>
</tr>
</thead>
</table>
| 4 | **FSPs should, where appropriate, allow customers to access, download, transfer and/or permit third parties to manage data about them**
|   | - **Accessibility:** FSPs should allow customers to download data about them in machine-readable format or through standardized APIs, depending on the FSP stage of development.
|   | - **Third-party permissions:** Customers should give third parties permission to download their data. |
| 5 | **FSPs should be held responsible and accountable for violation of customers’ data privacy**
|   | - **Liability:** A clear liability framework should be in place that ensures the responsible party is held accountable for data misuse and for harms caused by breaches of its respective data duties of care.
|   | - **Traceability:** FSPs need to be able to identify where data were improperly used. |
| 6 | **FSPs should be able to comprehensively test, validate and explain their use of data analytics or algorithms and models to customers**
|   | - **Justification:** Customers should have the right to request why a decision was made (e.g. why the model methodology is appropriate, why the output is justified).
|   | - **Challenge:** Customers should have the right to correct incorrect or incomplete data about them held by a FSP. |
Implications for Financial Regulators

### Licensing

At licensing stage key Non-Traditional Data risks to be unpacked including:
- Screen-scraping
- Algorithms and decisioning
- Cybersecurity risk
- Consumer education
- Transparency

### Supervision

As part of our risk-based supervisory activities **new risks associated with the use of Non-Traditional Data to be inspected** including:
- Discrimination
- Disclosure framework
- Dispute mechanism
- Data misuse
- Data breaches and privacy

### Enforcement

As part of enforcement and investigation activities the following will be required:
- Data Management/ APIs standards
- Digital audit trails/cybersecurity detection
- Specific disclosure mechanisms e.g. possible warnings.
- Misconduct analysis/identity management

### Consumer Education

- Owing to the complexity of some Non-Traditional Data value chain ensure consumer education plays a central role in demystifying and equipping consumers to make informed choices.

### Sandbox

- The sandbox as a facility to take a closer look at Fintechs using Non-Traditional Data in order to understand risks (and benefits) and that will inform how we supervise and regulate this activity.

### Regulatory Frameworks

- Derived from research, industry surveys, and data collected from the Sandbox, provide input to the CoFI Bill to ensure customers are treated fairly in usage of their data.