Data Protection Impact Assessment (DPIA)

Last updated 20-July-2020

Purpose of this Document

This model Data Protection Impact Assessment (DPIA) identifies a broad range of risks associated with the product offering of FutureFlow Research, Inc. (FutureFlow). In particular this DPIA aims to:

- describe the nature, scope, context and purposes of FutureFlow's processing
- assess necessity, proportionality and compliance measures
- identify and assess risks to individuals

This document serves as a template from which a more focused DPIA can be derived for a specific implementation of the FutureFlow platform. It details the control measures and mitigations that need to be put in place by FutureFlow and its partners to reduce and manage these risks.

This document applies to all personal data processed by FutureFlow in the course of its business (transaction monitoring and forensic analytics). This document may also be used by external stakeholders (including clients, Regulators and other third parties) seeking information about how FutureFlow processes personal data.

FutureFlow aims to embed good data protection practice in the working culture of its staff and operations.

Table of Contents

Identifying the need for a DPIA

FutureFlow's Transaction Monitoring and Forensic Analytics Platform monitors the flow of funds in the financial system. The platform enables financial institutions to contribute pseudonymized transactional data in bulk to a central Anti-Money-Laundering Utility Platform at a pre-suspicion level. The Utility enables multiple financial institutions, Regulators, and agencies to work together to detect and ultimately tackle Electronic Financial Crime.

Electronic Financial Crime poses a critical challenge to financial infrastructure and inflicts an enormous social and economic toll on the lives it touches. The problem has proven difficult to tackle due to, among other challenges, the inability of multiple financial institutions and agencies to collaborate effectively in fighting what is essentially a distributed, networked phenomenon. FutureFlow's collaborative, Utility-centric approach to detecting financial crime through cross-bank data processing and analytics opens the prospect of higher detection rates with lower false positives, while reducing the burden of scrutiny on each individual and business consumer.

As this processing relies on relatively new and innovative technologies (network analytics and machine learning) and involves pooling data from multiple sources, it appears as something that may pose high risk to individuals. As such, it requires a DPIA to be completed.

Description of the Processing

The nature of the processing

Pre-processing stage

FutureFlow can operate as a cross-bank Anti-Money-Laundering Utility in two modes:

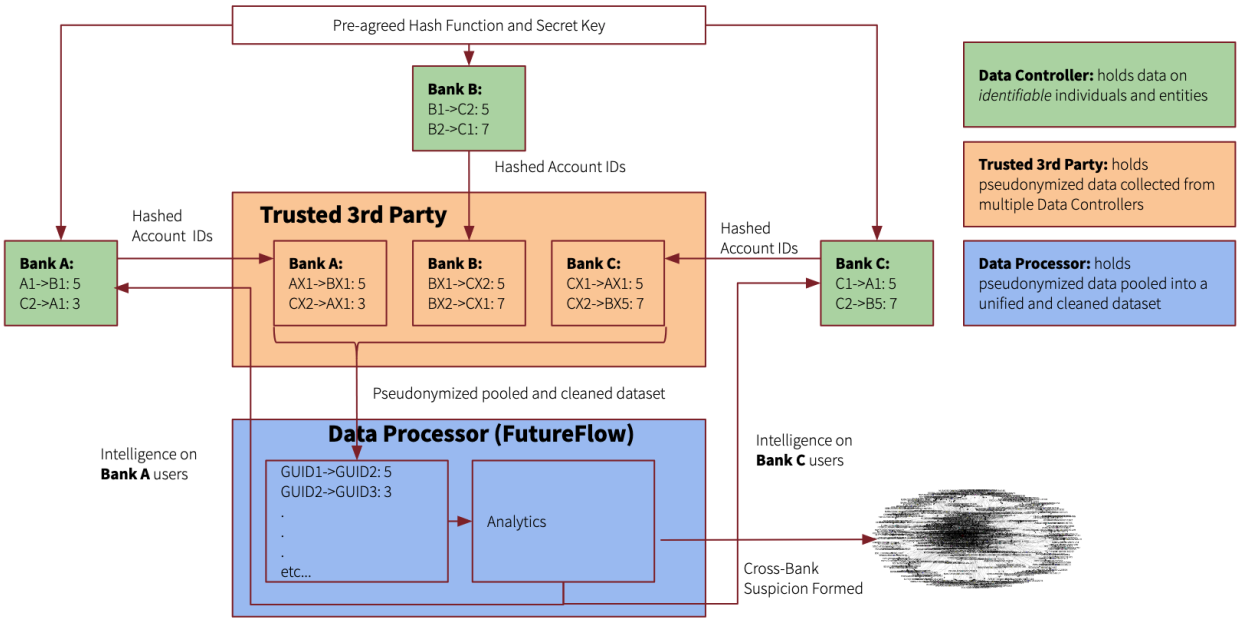
- **Indirect Mode:** FutureFlow operates as Data Processor, receiving transactional data with pseudonymized account identifiers from a Trusted Third Party which first receives pseudonymized transactional data from Data Controllers
- **Direct Mode:** FutureFlow operates as Data Processor, receiving transactional data with pseudonymized account identifiers directly from Data Controllers

While the core of FutureFlow's processing remains the same, the two modes differ in how the data on which FutureFlow operates is first pre-processed. While the Indirect Mode introduces additional coordination challenges, it offers a more meaningful level of obfuscation, thus reducing the level of risk to individuals.

Indirect Mode

In the Indirect Mode a Trusted Third Party facilitates the exchange of data between each feeder financial institution and the Central Utility (FutureFlow). A representative profile and the nature of business of a suitable Trusted Third Party is a large Consultancy, a System Integrator, or another organization of a similar statute that commands a trusting business relationship with multiple Data Controllers by the

nature of its business. For example, in the first pilot of FutureFlow on real-life data, a Big-4 Consultancy served as a de-facto Trusted Third Party.



In the Indirect Mode, the Trusted Third Party helps multiple banks to coordinate and agree on a common convention for identifying and pseudonymizing account identifiers in the transactional datasets, which will subsequently be processed by FutureFlow. As described in the above diagram, the process includes:

1. Agreeing a common convention on identifying the sending and receiving account in a transaction (such as using an IBAN, a combination of Account Number and Sort Code, etc.)
2. Agreeing a common Hash Function for pseudonymization. The chosen Hash Function should be of industry-recognized strength and should conform to the standard principles of hashing:
    a. **One-way:** a hash is irreversible to original value computationally (only by trial and error)
    b. **Consistent:** hashing the same value always produces the same hash
    c. **Collision-Free:** hashing two different values to the same hash is highly unlikely
3. Agreeing a common Secret Key and a common convention of mixing this Secret Key with the account identifier agreed in Step 1 above (NOTE: the Trusted Third Party should be unaware of the Secret Key, and a sufficiently complex Secret Key should be agreed by the banks to reduce the risk of re-identification at the Trusted Third Party level)

Each participating bank submits its transactional dataset, with account identifiers pseudonymized according to the steps outlined above, to the Trusted Third Party. The Trusted Third Party then combines multiple datasets into a pooled dataset and performs the following transformations:

1. **Deduplication**: for example, a *debit* transaction in Bank A's dataset, in which a Bank A account *sends* funds to a Bank B account, may also be submitted as a *credit* transaction in Bank B's dataset, where the Bank B account *receives* the funds from the Bank A account. After deduplication, the two transactions from the two separate datasets are represented by one transaction in the pooled dataset where Bank A account sends funds to the Bank B account.
2. **Further Obfuscation (optional)**: the account identifiers in the pooled dataset are pseudonymized by banks as described above. By pre-agreeing the Hash Function, the Secret Key,

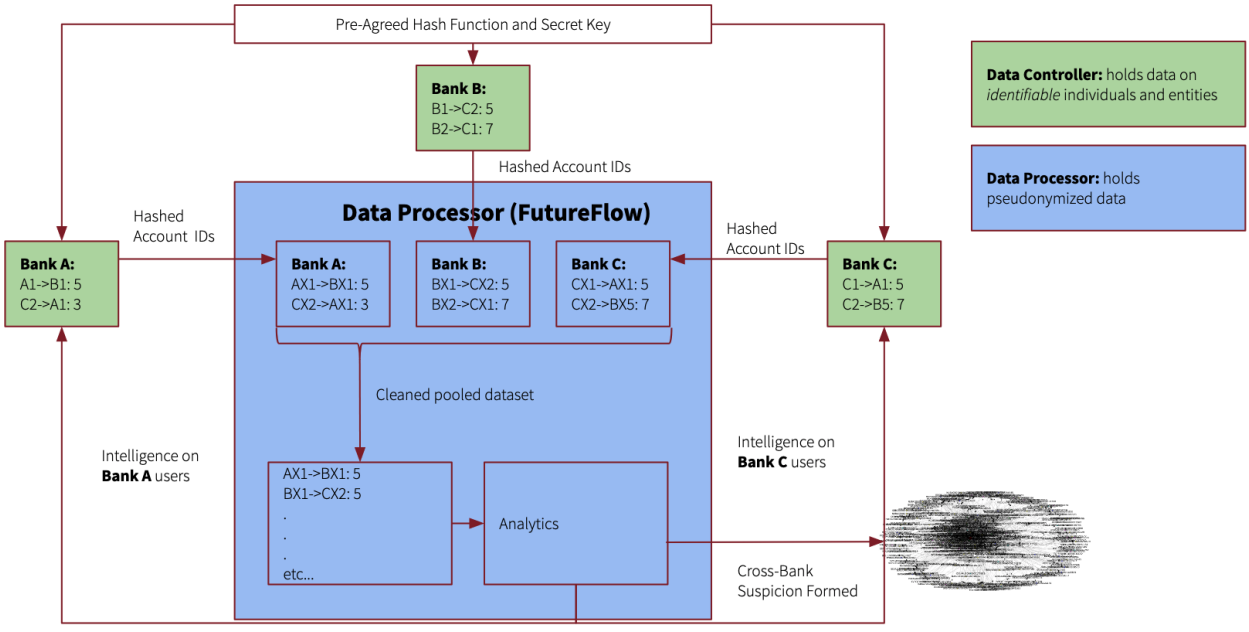
and the identifier structure, the participating banks enable the Trusted Third Party to match the same account listed in multiple banks' datasets by its hash without seeing it in plain sight. This enables the Trusted Third Party to replace each account hash in the pooled dataset with a random identifier, since this can be done consistently for each account across the whole pooled dataset. This optional step introduces even further obfuscation of the original account identifier in the final pooled dataset that is sent to FutureFlow for processing, since no backward computational connection remains between the random identifier in the final pooled dataset and the pseudonymized hash that it replaced in the original pooled dataset.

Summary of transformation steps performed on an account identifier in the diagram above:

- In plain sight: A1 hypothetical account identifier at Bank A
- Pseudonymized: AX1 hash of A1, assuming HashFunc(SecretKey + A1) = AX1
- Further Obfuscated (optional): GUID1 random GUID (Globally Unique Identifier) or similar

Direct Mode

In the Direct Mode, each feeder financial institution submits the transactional data with pseudonymized entity identifiers directly to the Central Utility (FutureFlow). In this case, the participating financial institutions agree among themselves on the same synchronization and pseudonymization techniques as in the Indirect Mode, while FutureFlow performs deduplication and other cleaning operations on the pooled dataset, without relying on the Trusted Third Party.



While the Direct Mode reduces the level of complexity in coordination and data exchanges between FutureFlow and the participating financial institutions, it offers a lower degree of data obfuscation compared with the Indirect Mode, since FutureFlow operates on pseudonymized data directly.

Processing stage

At the data processing stage, FutureFlow performs synthetic tokenization of transaction amounts in the underlying dataset, followed by bootstrapping and topological analysis of networks that may result from tracing the histories of the individual amount tokens. This “follow the money” approach maps out complex non-linear cross-bank account relationships, extending each participating bank’s field of visibility beyond its own boundaries.

This novel approach to cross-bank financial network mapping and analytics enables multiple banks to tackle financial crime jointly, as opposed to individually, and using only their own data. It prevents malicious actors from deliberately moving their funds among multiple financial institutions to avoid detection. It empowers financial institutions with a more holistic and in-depth view of the movement of funds across the entire financial system, thus spotting at a system-level unusual behaviors and transaction patterns that may constitute financial crime.

FutureFlow enables two complementary approaches for generating intelligence from its processing:
Reactive Approach: this approach relies on the leading intelligence that can be provided by the financial institutions to generate alerts and insights. For example, some financial institutions may choose to submit transaction flags, seed accounts, or other types of leading intelligence that FutureFlow can use to highlight the relevant parts of the underlying account universe as being strongly associated with the provided accounts.
Proactive Approach: this approach offers automated lead generation, without relying on any leading intelligence supplied by the banks. In this approach, FutureFlow automatically evaluates and ranks the complexity of the generated networks, highlighting those that may deserve particular attention by the Financial Crime analysts of the participating banks.

The Reactive and Proactive approaches are complementary and self-reinforcing in enabling the participating financial institutions to conduct transaction monitoring, forensic analytics, and continuous process improvement of the underlying account base. At a minimum, they enable the following use-cases

- Cross-bank alert generation: enabling blind and automatic cooperation and information sharing regarding problematic accounts across the participating banks
- Case triage: enabling each bank to triage alerts and concerns with the benefit of wider cross-bank intelligence
- Lead generation: bringing each participating banks’ attention to some problem areas in the underlying account base that may have never been discovered before

The processing described above is performed at a “pre-suspicion” stage, meaning that financial institutions can submit their transactional data without necessarily having any preconceived knowledge or suspicion of any integrity risks present across the transacting accounts in the submitted dataset. The purpose of the processing is to empower multiple financial institutions to collectively spot, assess, and report criminal transacting patterns by understanding the big picture of the flow of funds. This is discussed in greater detail in the Purposes of the Processing section below.

In order to comply with article 30 of the GDPR FutureFlow will record their processing in their Record of Processing Activity (ROPA).

The scope of the processing

Through its activities as Data Processor, FutureFlow operates on the following categories of personal data in a pseudonymized format:

- Account identifier (i.e. account number + sort code/routing number combination, IBAN, etc.)
- Transaction value
- Transaction IDs
- Time-stamp
- Flags (optional)

FutureFlow does not process Special Category Personal Data as defined by article 9 of the GDPR.
FutureFlow only processes the data provided by the Data Controller clients. The amount of data supplied, the number of data subjects affected, and the geographical scope of the data is likely to vary on a project to project basis, as these specifics depend on FutureFlow's clients.

In ad-hoc projects where the processing takes place once and is not intended to be repeated or supplemented by new data on a regular basis, FutureFlow retains the data for as long as necessary to complete the synthetic tokenization, the network mapping, and the subsequent analysis, which is likely to be in the range of few weeks to 2-3 months.

In prolonged ongoing projects, where the processing is intended to be done repeatedly on the incrementally submitted fresh data, the data retention policy for each submission batch depends on the scope agreed with the client Data Controller(s) supplying the data. Given the nature of the processing, the more recently submitted data makes the previously submitted data incrementally less relevant, but not entirely irrelevant. Therefore, the data retention and the data lineage specifications form the essential part of the project requirements.

The context of the processing

FutureFlow does not maintain a direct relationship with the entities and the individuals whose transactional data it processes. Furthermore, since the data provided to FutureFlow is limited in scope and is pseudonymized to such a degree that it is rendered effectively anonymous, FutureFlow is unable to communicate directly with any individuals in the underlying transactional data. For more information about data preprocessing and pseudonymization, see the Pre-Processing Stage portion of The Nature of the Processing section above.

Given the novelty of the FutureFlow use-case, it is assumed that data subjects do not know that their personal data is being processed by FutureFlow and further that they have no reasonable expectation for their data to be processed specifically by FutureFlow. However, it is also assumed that data subjects recognize that financial and economic crime is a significant threat that financial institutions are required to mitigate, and that their financial institution, acting as Data Controller, routinely processes their personal data for purposes of financial crime prevention and monitoring, including by engaging various Data Processors, such as FutureFlow.

FutureFlow is not currently signed up to any approved codes of conduct or certification schemes.

The purposes of the processing

As previously described, the ultimate purpose of the processing is to empower multiple financial institutions to collectively spot, assess, and report potentially criminal accounts and transacting patterns by understanding the big picture of the flow of funds. By conducting Proactive and Reactive analysis, as described above, on the pooled dataset from multiple banks, FutureFlow flags accounts and transactions that it considers potentially suspicious and submits the flags on such accounts to each respective financial institution. These flags can be used by the financial institution as a form of intelligence to investigate the account or transaction further, or as a basis for generating a Suspicious Activity Report.

Furthermore, the analysis may often flag shared cross-bank topologies that involve several accounts from multiple banks. While technically such patterns may represent a relevant form of intelligence, at this stage of the processing it is not feasible to share such information with each financial institution whose accounts are included in the topology, since the information includes accounts from other institutions. However, such intelligence can serve as a basis for generating a cross-bank 'case' of suspicion, which may identify a clear threat to the integrity of the financial system. The formation of the case can serve as the facilitator of a much more comprehensive collaboration among the participating financial institutions regarding the set of accounts included in the case, which will not involve FutureFlow and is therefore outside of the scope of this DPIA. For more details on this, see the Necessity and Proportionality section of this document below.

Consultation Process

FutureFlow has invested over three years into the ongoing dialog and consultations with various stakeholders in the Financial Crime Prevention industry, including regulators, financial institutions, and security/technology specialists. Among other avenues, these interactions took place in the UK market through FutureFlow's formal participation in:

- The ICO Regulatory Sandbox (2019-2020)
- The FCA Regulatory Sandbox (2018)
- The FCA Anti-Financial-Crime TechSprint (2019)
- The Accenture FinTech Innovation Lab Accelerator (2019)

Additionally, FutureFlow conducts similar consultations with equivalent stakeholders in other European jurisdictions covered by GDPR, including Sweden, Netherlands, and Finland.

Through these interactions, FutureFlow has been continuously tailoring its platform to cater to the following unmet needs of the industry and to mitigate the following shortcomings of some existing technologies:

- A technology-centric systematic approach to collaboration and information sharing across financial institutions that can enable a holistic customer view
- An approach to collaboration and information sharing that can take place before a suspicion has been flagged

Risks and limitations

- Gathering large volumes of personal data in-house in order to assemble a holistic customer view can pose more danger to individual privacy and lead to greater damages in case of data breaches
- Frontier technologies that enable computation on encrypted data across multiple institutions, such as homomorphic encryption, are not yet ready for generic deployment and broad use-cases
- Traditional Supervised Machine Learning lacks clear explainability and requires vast amounts of reliable labeled data, which is often unavailable in the Financial Crime space

Considering the highly specialized and technical nature of the processing, FutureFlow considers it inappropriate to seek individuals' consultations on this subject, as they are unlikely to reflect on the realities and the complexities of the underlying business problem. Moreover, since each member of the Financial Crime Prevention ecosystem described above is also an individual, inevitably these consultations often involve the individuals' personal reflections and impressions as Data Subjects, as opposed to just field professionals. At this stage, FutureFlow is not planning to seek any specific individual consultations on its processing.

Assessment of the Necessity and Proportionality

Background

As a Data Processor, FutureFlow is not responsible for selecting an appropriate legal basis for processing on behalf of the Data Controller clients that submit their data. However, FutureFlow sees articles 6,1(F) and 6,1(C) of the GRPD as two relevant bases for its clients to consider while incorporating FutureFlow into their Anti Financial Crime technology capabilities.

Ongoing monitoring and analysis of own accounts

For the purposes of the ongoing monitoring and analysis of a broad range of accounts, FutureFlow sees article 6,1(F) of the GDPR as the most appropriate legal basis for its clients to use. This article stipulates that the processing should be 'necessary for the purposes of the legitimate interests pursued by the controller or by a third party, except where such interests are overridden by the interests of the fundamental rights and freedoms of the data subject which require protection of personal data, in particular where the data subject is a child.'

Clients seeking to rely on article 6,1(F) as basis for processing will be required to complete their own legitimate interests assessment (LIA), as detailed on the Information Commissioner's Office's (ICO's) website. FutureFlow's analytics platform has been designed to utilize the least amount of personal data, which is pseudonymized to a degree that renders it effectively anonymous. This should help clients choosing to involve FutureFlow as Data Processor to pursue the legitimate interest of monitoring and preventing financial crime to fulfill the LIA requirements.

Focused analysis of suspicious cross-bank relationships

Note: the processing discussed in this section is separate from the FutureFlow processing and is not covered by this DPIA. It may take place optionally at the discretion of Data Controllers as a result of the FutureFlow processing.

In cases where the FutureFlow processing reveals a suspicious cross-bank cluster involving accounts of multiple banks, FutureFlow's clients may wish to collaborate more closely with each other on investigating and reporting the cross-bank case jointly. In such instances, FutureFlow's clients may choose to rely on article 6,1(C) as their basis for processing of personal data on these specific accounts further. This condition states that processing should be 'necessary for compliance with a legal obligation to which the controller is subject'.

Typically, anti-money laundering and related legislation *requires* financial institutions to investigate instances of financial and economic crime once a Suspicious Activity Report (SAR) has been raised. Furthermore, the more recent legislation *allows* financial institutions to collaborate and to share information with each other once a SAR has been raised, or to raise super-SARs jointly.

With the above in mind, FutureFlow's clients will need to consider whether the cross-bank topologies generated and flagged by the FutureFlow platform represent a sufficient body of evidence for raising a SAR (or a super-SAR) and whether the institutions need to be conducting a joint review of the accounts by relying on article 6.1(C) as basis for processing for sharing any further information that may be required. For avoidance of doubt, this further information sharing, if any, will not involve FutureFlow as Data Processor and are not covered by this DPIA.

Concluding notes

FutureFlow's analytics platform provides financial institutions with a viable and secure means to share their transactional data and to gain insights at a 'pre-suspicion' level, to target and address the integrity risks more effectively, to raise more robust, evidence-based SARs, and to cooperate more closely post-SAR submission. Through the ongoing research described in the Consultation Process section above, FutureFlow concludes that there are currently few, if any, alternative technology offerings that enable financial institutions to gain the same level of insight into their data, with a lower degree of intrusiveness towards their data subjects.

As described above, the FutureFlow analytics aims to help financial institutions (data controllers) to expand their existing financial crime intelligence (Reactive Approach) and to generate new intelligence automatically (Proactive Approach). In both approaches, the relevant flags are submitted back to the participating institutions (data controllers) for further analysis and investigations. In order to prevent data controllers using these flags for other purposes, such as immediately denying individuals suspected of financial crime banking services (i.e. to avoid function creep), FutureFlow's contract with its data controllers shall state that all insights provided by FutureFlow should be used as an intelligence source only, and that no immediate action should be taken against account holders/data subjects as a result of FutureFlow flagging their account.

FutureFlow's processing does not appear to trigger article 10 of the GDPR, because the processing takes place before the providing data controller has a reasonable suspicion that a specific offence has taken place (i.e. the processing takes place pre-suspicion). However, it is likely that data controllers that rely on FutureFlow's insights as part of their evidence when investigating and enforcing against data subjects suspected of financial crime would trigger article 10. Consequently, these data controllers need to ensure that they select a suitable condition to process the criminal conviction data from schedule 1 of the DPA18. It is likely that FutureFlow's clients (i.e. the data controllers) will look to rely on Schedule 1, condition 10 of the DPA18 (Preventing or detecting unlawful acts) to process GDPR article 10 data.

Since the data processed by FutureFlow is first pseudonymized by data controllers to such a degree that individual data subjects are no longer identifiable, it is not technologically possible for FutureFlow to fulfil data subject rights requests directly. Furthermore, as a processor of personal data, FutureFlow may not always be the appropriate party to assist data subjects in exercising their GDPR Article 12 rights. However, FutureFlow will wherever possible assist their clients (i.e. data controllers) in completing and complying with data subject rights requests and will communicate with data subjects in a concise, transparent, intelligible and easily accessible form and without undue delay.

Since FutureFlow receives data directly from data controllers, data subjects are unlikely to be aware that FutureFlow is holding or processing their personal data. In order to ensure that data subjects are aware that FutureFlow may be processing their personal data, FutureFlow clients are encouraged to include the necessary language in their privacy policies about how FutureFlow may process personal data.

FutureFlow is designed with the aim of storing and processing the data in its geographical origin. When deployed in the Cloud, FutureFlow deploys resources in Regions and Zones of the appropriate geography. At present, FutureFlow excludes from its architecture any Cloud functionality that involves any ambiguity as to the geographical location for data storage or processing. When deployed on-premises with client financial institutions (data controllers), FutureFlow relies on data controllers' own compliance with data residency requirements, as relates to the location of their compute infrastructure.

Identification and Assessment of Risks

FutureFlow's objective when providing financial analytics services is to support its clients in detecting and ultimately preventing instances of financial and other economic crime utilizing transactional data. FutureFlow and its clients must recognize that by utilizing these services, they are increasing their exposure to data protection risks, which must be considered and, where possible, mitigated.

The data protection risks FutureFlow and its clients need to consider can broadly be categorized as:

- Principle breaches of the GDPR (Article 5)
- Inability to fulfil GDPR data subject rights requests (Chapter 3: Articles 12-23)
- Risks associated with domestic and international data transfers (Article 32 and Articles 44-50)

Risks resulting from FutureFlow's processing could affect:

- Data subjects (i.e. account holders at FutureFlow's client banks).
- FutureFlow's Clients
- FutureFlow itself.

FutureFlow's clients will need to consider on a case-by-case basis whether they are willing to accept the risks highlighted in this Document. The below model risk assessment should be used as a reference point to demonstrate how FutureFlow effectively manages the risks associated with its processing. It needs to be considered in the context of a specific implementation of the FutureFlow platform and expanded or amended where necessary to reflect the specific circumstances of the implementation. Further information on FutureFlow's risk scoring methodology is contained in Annex A of this document.

Risk Ref.	**Owner:** Vadim Sobolevski	**Gross Risk**			
	Risk Description	**Likelihood**	**Impact**	**Total Score**	**Risk Status**
1	Customer data, including personal data, inappropriately submitted to FutureFlow, caused by incorrect understanding by Data Controllers of their obligations to their Data Subjects. This may result in unlawful data processing on behalf of FutureFlow.	2	4	8	High
	Existing Controls and Evidence	**Net Risk**			
	FutureFlow operates as Data Processor; therefore, it remains up to FutureFlow's clients, acting as Data Controllers, to choose the appropriate Basis for Processing before sending their data to FutureFlow. With this Model DPIA, FutureFlow has provided clear guidance to Data Controllers about which Basis for Processing is likely to be most appropriate for its use-cases. FutureFlow's clients are likely to be sophisticated financial institutions with legal departments capable of assessing this guidance.	**Likelihood**	**Impact**	**Total Score**	**Risk Status**
		1	2	2	Low

Risk Ref.	**Owner:** Vadim Sobolevski	**Gross Risk**			
	Risk Description	**Likelihood**	**Impact**	**Total Score**	**Risk Status**
2	Software code exhibits bias or targets specific segments of the population, caused by poor design choices or exposure to biased training data, resulting in unfair data processing.	2	4	8	High
	Existing Controls and Evidence	**Net Risk**			

	FutureFlow operates on transactional data where entity identifiers are obfuscated to such an extent as to be effectively anonymous. This makes it impossible for the system to knowingly target specific accounts for processing, or to knowingly exclude specific accounts from processing, except for reasons of data quality that are described below. Furthermore, since FutureFlow is based primarily on unsupervised models, it does not rely on training data that may potentially contain inherent bias or be mislabeled.	**Likelihood**	**Impact**	**Total Score**	**Risk Status**
		1	3	3	Low

Risk Ref.	**Owner:** Vadim Sobolevski	**Gross Risk**			
	Risk Description	**Likelihood**	**Impact**	**Total Score**	**Risk Status**
3	Data Subjects are not provided with significant information about how FutureFlow processes their data, caused by lack of disclosure on behalf of Data Controllers, resulting in un-transparent data processing by FutureFlow.	3	3	9	High
	Existing Controls and Evidence	**Net Risk**			
	FutureFlow has created a Transparency Statement, which it provides to each client Data Controller. It is up to each Data Controller to use this Statement to communicate to its Data Subjects that their data is processed by FutureFlow. Furthermore, as stated in the FutureFlow Data Protection Policy, the company has a commitment to communicate with Data Subjects (its clients' customers) in a clear and transparent manner, and where feasible, to communicate Data Subject Rights Requests to Data Controllers.	**Likelihood**	**Impact**	**Total Score**	**Risk Status**
		1	2	2	Low

Risk Ref.	**Owner:** Vadim Sobolevski	**Gross Risk**			
	Risk Description	**Likelihood**	**Impact**	**Total Score**	**Risk Status**

4	Data submitted to FutureFlow is processed for purposes other than those pre-agreed with Data Controllers, resulting from deliberate application of software code designed for use-cases other than those pre-agreed with Data Controllers. This may result in FutureFlow processing personal data illegitimately.	1	4	4	Medium
	Existing Controls, Evidence, and Notes	**Net Risk**			
	In on-premises deployments, Data Controllers (or Trusted Third Party) remain in charge of the physical infrastructure where the data resides and where the processing takes place. While it's impractical for FutureFlow to disclose every single element of its code (particularly compiled code) to Data Controllers/Trusted Third Party before the processing, generally the controllers are at least aware of the output of the processing, as well as the data egress/ingress to/from the environment. This makes it very difficult to process the data for illegitimate purposes without the controller's awareness. Note: Future single-client Cloud deployments (involving just one Data Controller) will likely be executed under the Data Controller's own subscription, so in practice they are no different from above. Note: Future Multi-tenant Cloud deployments (involving multiple Data Controllers) will likely be executed under a neutral utility subscription, making it difficult for each individual Data Controller to be fully aware of how their data is used by Data Processor(s), including FutureFlow. However, such deployments are likely to be subject to multi-institution governance and oversight. Moreover, in	**Likelihood**	**Impact**	**Total Score**	**Risk Status**
		1	2	2	Low

	such deployments the FutureFlow platform is likely to be just an element within a broader architecture, making it impossible to somehow be using the data for illegitimate purposes.				

Risk Ref.	**Owner:** Vadim Sobolevski	**Gross Risk**			
	Risk Description	**Likelihood**	**Impact**	**Total Score**	**Risk Status**
5	FutureFlow clients may send FutureFlow more data than is strictly necessary to complete the processing, caused by poor communication or improper understanding of what data is necessary for FutureFlow to perform the processing. This may result in FutureFlow receiving irrelevant data, or more data than necessary, particularly in cases where Data Minimization principles are needed to be followed to perform the processing under Legitimate Interest as Basis for Processing.	2	2	4	Medium
	Existing Controls and Evidence	**Net Risk**			
	FutureFlow was designed according to Data Minimization principles and uses only the minimum information necessary to trace the flow of funds from one account to another. This data is to be supplied by Data Controller(s) according to a pre-agreed schema, and via pre-agreed channels, which eliminates the risk of receiving more data than necessary. In cases where Data Processors, either accidentally or deliberately, submit data to FutureFlow outside of the pre-agreed channels and that data does not conform to the pre-agreed schema (i.e. contains more information than necessary) such data by definition would not be deemed fit for processing. Such data would be destroyed.	**Likelihood**	**Impact**	**Total Score**	**Risk Status**
		1	1	1	Low

Risk Ref.	Owner: Vadim Sobolevski	Gross Risk			
	Risk Description	**Likelihood**	**Impact**	**Total Score**	**Risk Status**
6	FutureFlow clients may send inadequate data for processing, resulting from poor internal data quality standards. This may lead to FutureFlow operating on inadequate data.	3	2	6	Medium
	Existing Controls and Evidence	**Net Risk**			
	In some instances, transactional data may indeed be inadequate by nature. For example, a Credit Card or a Contactless Card transaction may have a missing recipient identifier. The platform is designed to be aware of such limitations. Furthermore, the platform is designed to operate under severe informational gaps (i.e. missing a complete picture of the flow of funds). This means that eliminating transactions with inadequate information (i.e. missing account identifiers/etc., now transaction amounts, etc.) should not result in complete invalidation of the produced output)	**Likelihood**	**Impact**	**Total Score**	**Risk Status**
		3	1	3	Medium

Risk Ref.	Owner: Vadim Sobolevski	Gross Risk			
	Risk Description	**Likelihood**	**Impact**	**Total Score**	**Risk Status**
7	FutureFlow clients sending inadequate data for processing, resulting from lack of coordination among Data Controllers in the process of pseudonymization. This may lead to FutureFlow operating on inadequate data and producing wrong linkages.	3	5	15	High
	Existing Controls, Evidence, and Notes	**Net Risk**			
	Data quality and coordination are key in achieving the right result in the process of pseudonymization across multiple	**Likelihood**	**Impact**	**Total Score**	**Risk Status**
		2	1	2	Low

	parties, as even minor mistakes or misunderstandings may result in a complete breakdown of hash-based linking. This risk is unlikely lead to any harm to specific Data Subjects, for example via improperly linking an incorrectly identified benign account with a malicious account. This is due to the Collision-Free property of industry-standard hash functions (i.e. it is virtually impossible for a hash of one incorrectly articulated account identifier to result in an exact hash value of some other correctly or incorrectly articulated account identifier). However, this risk is likely to result in the output produced by FutureFlow being irrelevant and not useful for clients. With the Indirect Mode of preprocessing, the Trusted Third Party is responsible for coordination, aggregation, and de-duplication across multiple Data Controllers and must take the necessary steps to address this risk. With the Direct Mode of processing, FutureFlow is responsible for addressing this risk. This risk is treated through a combination of Coordinating Actions and Data Quality check. Coordinating Actions: prior to the processing, Data Controllers are to be educated on how to properly articulate an account identifier, what hash function to choose, how to correctly agree on the Common Secret Key, and what format/encoding the original identifier needs to be in. This should involve working out common misunderstandings, such as removing trailing zeros from account identifiers or coordinating on exact formats of the commonly confused characters such as				

	'-' vs. '—' that may appear similar to a human, but may be encoded differently by computer systems. **Data Quality Checks:** a small sample of designated test account identifiers may be hashed by all Data Controllers and checked centrally by Data Processor to ensure that the hashing was done properly and that the same identifiers hashed by different Data Controllers indeed result in the same hashes. Once full datasets have been submitted, Trusted Third Party or Data Processor may perform a broad linkage test, to determine what portion of the data links across various Data Processors' silos. Any mistakes are likely to result in a complete breakdown of cross-silo linkages and should therefore be evident.				

Risk Ref.	**Owner:** Vadim Sobolevski	**Gross Risk**			
	Risk Description	**Likelihood**	**Impact**	**Total Score**	**Risk Status**
8	Data processed by FutureFlow is inaccurate resulting in the end point insights (flags) created by FutureFlow's systems to be inaccurate. This may lead to Suspicious Activity Reports being made wrongly.	2	4	8	High
	Existing Controls and Evidence	**Net Risk**			
	The data submitted to FF for analysis is likely to be from banks and other financial institutions where data is, by necessity and by virtue of other regulations, highly accurate. Also, as the data submitted to FutureFlow will usually be the combined data of several organisations, any inaccuracies would be flagged when the data is combined pre analysis (i.e. both bank A and bank B would submit information about an individual transaction and if they had	**Likelihood**	**Impact**	**Total Score**	**Risk Status**
		1	2	2	Low

	different values noted for the transaction this would be flagged before the data is deduplicated).				

Risk Ref.	**Owner:** Vadim Sobolevski	**Gross Risk**			
	Risk Description	**Likelihood**	**Impact**	**Total Score**	**Risk Status**
9	Re-identification of pseudonymized account identifiers in the data submitted to FutureFlow due to hash reverse engineering. This may result in re-identification of data subjects.	1	3	3	Medium
	Existing Controls, Evidence, and Notes	**Net Risk**			
	By choosing a sufficiently strong, industry-standard hash function, Data Controllers would be effectively eliminating the risk of a computational reverse engineering for the foreseeable future. Once the computing industry comes closer to widespread commercial adoption of Quantum Computing or a similar supercomputing functionality, this Risk will need to be reviewed in accordance with the new reality. Note: Attempts at reverse engineering by FutureFlow while data remains only in possession of FutureFlow (i.e. there are no Data Breaches) is equivalent to Illegitimate Processing – see Risk 4 above. Note: Attempts at reverse engineering by a third party is equivalent to a Data Breach. See Data Breach-related Risks below. Note: even a successful reverse engineering of a hash stored in FutureFlow would only reveal the fact that a certain account identifier exists at a certain financial institution. It would not reveal any significant personal information relating to that account, such	**Likelihood**	**Impact**	**Total Score**	**Risk Status**
		1	2	2	Low

	as an individual's name, address, age, etc., because no such information is ever sent to FutureFlow.				

Risk Ref.	**Owner:** Vadim Sobolevski	**Gross Risk**			
	Risk Description	**Likelihood**	**Impact**	**Total Score**	**Risk Status**
10	Re-identification of pseudonymized account identifiers in the data submitted to FutureFlow via a Rainbow Attack. This may result in re-identification of data subjects.	1	3	6	Medium
	Existing Controls, Evidence	**Net Risk**			
	A Rainbow Attack is a repeated trial-and-error attempt to match various inputs to an output in order to guess empirically the plain-text equivalent of a hash. While it is trivial to perform a Rainbow Attack to any account identifier, Data Controllers submitting their data to FutureFlow would significantly reduce this risk by using a sufficiently long and complex Secret Key. Note: Attempts at a Rainbow Attack by FutureFlow while data remains only in possession of FutureFlow (i.e. there are no Data Breaches) is equivalent to Illegitimate Processing – see Risk 4 above. Note: Attempts at a Rainbow Attack by a third party is equivalent to a Data Breach. See Data Breach-related Risks below. Note: even a successful Rainbow Attack on a hash stored in FutureFlow would only reveal the fact that a certain account identifier exists at a certain financial institution. It would not reveal any significant personal information relating to that account, such as an individual's name, address, age, etc., because no such information is ever sent to	**Likelihood**	**Impact**	**Total Score**	**Risk Status**
		1	2	2	Low

	FutureFlow.				

Risk Ref.	**Owner:** Vadim Sobolevski	**Gross Risk**			
	Risk Description	**Likelihood**	**Impact**	**Total Score**	**Risk Status**
11	Re-identification of pseudonymized account identifiers in the data submitted to FutureFlow resulting from a disclosure of one Data Controller's dataset to another. This may result in re-identification of data subjects.	2	4	8	High
	Existing Controls, Evidence, and Notes	**Net Risk**			
	As discussed above, the risk of a Rainbow Attack on pseudonymized account identifiers is significantly reduced in general cases by choosing a sufficiently long and complex Secret Key, which is known only to Data Controllers. However, in the specific case where one Data Controller's data is disclosed to another, the Secret Key IS known by the recipient Data Controller, so a Rainbow Attack becomes trivial regardless of the complexity of the Secret Key. With the Indirect Mode of preprocessing, this risk exists at the Trusted Third Party level and should be addressed there. In the Direct Mode of processing this risk resides with FutureFlow. Note: Sending any data received from Data Processors outside of FutureFlow violates the Security section of the FutureFlow Data Protection Policy. This includes sending one Data Controller's data to another Data Controller. Furthermore, the Staff Training section of the FutureFlow Data Protection Policy ensures that all staff with any exposure to Data Controller's data is educated on the specifics of hashing and its side effects. Note: sending data back to Data	**Likelihood**	**Impact**	**Total Score**	**Risk Status**
		1	4	4	Medium

	Controller is also addressed in Risk 5 above. Note: even a successful Rainbow Attack on or a reverse engineering attempt of a hash stored in FutureFlow would only reveal the fact that a certain account identifier exists at a certain financial institution. It would not reveal any significant personal information relating to that account, such as an individual's name, address, age, etc., because no such information is ever sent to FutureFlow.				

Risk Ref.	**Owner:** Vadim Sobolevski	**Gross Risk**			
	Risk Description	**Likelihood**	**Impact**	**Total Score**	**Risk Status**
12	Data subjects covered in the data processed by FutureFlow are re-identified, caused by the FutureFlow data being linked with external datasets. This may result in loss of privacy by Data Subjects.	2	4	8	High
	Existing Controls, Evidence, and Notes	**Net Risk**			
	Note: Attempts at reverse engineering by FutureFlow while data remains only in possession of FutureFlow (i.e. there are no Data Breaches) is equivalent to Illegitimate Processing – see Risk 4 above. Note: Inappropriate linking by FutureFlow while data remains only in possession of FutureFlow (i.e. there are no Data Breaches) is equivalent to Illegitimate Processing – see Risk 4 above. Note: Attempts at Linking by a third party is equivalent to a Data Breach. See Data Breach-related Risks below. Note: Linking FutureFlow data with other	**Likelihood** 1	**Impact** 2	**Total Score** 2	**Risk Status** Low

	data for purposes of re-identification would normally require some usable key-like reference in the data on which the linking can be done. The only candidate for this is the pseudonymized account identifier. Therefore, a successful Rainbow Attack or hash reverse-engineering has to take place first (these Risks are addressed above). Beyond this the linking is not feasible.				

Risk Ref.	**Owner:** Vadim Sobolevski	**Gross Risk**			
	Risk Description	**Likelihood**	**Impact**	**Total Score**	**Risk Status**
13	Data subjects covered in the data processed by FutureFlow are re-identified, caused by deliberate picking out of rare occurrence transactions in the data. This may result in loss of privacy by Data Subjects.	2	4	8	High
	Existing Controls, Evidence, and Notes	**Net Risk**			
	Rare occurrence transactions, such as infrequent low-value payments in rural areas, would normally require additional data to be discoverable (such as geolocation data, device IP addresses, etc.). Such data is not supplied to FutureFlow. Note: re-identification through other rare events, such as a large lottery win by an individual, while data remains only in possession of FutureFlow (i.e. there are no Data Breaches) is equivalent to Illegitimate Processing – see Risk 4 above. Note: attempts at re-identification through rare events by a third party is equivalent to a Data Breach. See Data Breach-related Risks below.	**Likelihood**	**Impact**	**Total Score**	**Risk Status**
		1	3	3	Low

<table>
<tr><td rowspan="2">Risk Ref.</td><td>Owner: Vadim Sobolevski</td><td colspan="4">Gross Risk</td></tr>
<tr><td>Risk Description</td><td>Likelihood</td><td>Impact</td><td>Total Score</td><td>Risk Status</td></tr>
<tr><td rowspan="3">14</td><td>Data leakage, caused by a security breach, resulting in potential loss of privacy by Data Subjects whose data is submitted to FutureFlow</td><td>3</td><td>4</td><td>12</td><td>High</td></tr>
<tr><td>Existing Controls, Evidence, and Notes</td><td colspan="4">Net Risk</td></tr>
<tr><td rowspan="2">In on-premises deployments, Data Controllers (or Trusted Third Party) remain in charge of the physical infrastructure where the data resides and where the processing takes place. In these instances, Data Controllers remain in charge of security and can exercise industry-standard measures to mitigate the risk of security breaches, such as limiting data ingress/egress and only allowing internal traffic into the infrastructure.

Note: Data Minimization and pseudonymization – two key principles of the FutureFlow design – ensure that in case of a Breach the leaked data should not render itself useful for illegitimate processing aimed at Data Subject re-identification, such as Rainbow Attacks, hash reverse engineering, and data linking.

Note: In planned Cloud deployments the FutureFlow platform will be part of a larger technical architecture and the processing will be subject to multi-institution governance. In such deployments, Data Security considerations will be delegated to the oversight bodies and the Cloud provider.</td><td>Likelihood</td><td>Impact</td><td>Total Score</td><td>Risk Status</td></tr>
<tr><td>1</td><td>3</td><td>3</td><td>Medium</td></tr>
</table>

<table>
<tr><td rowspan="2">Risk Ref.</td><td>Owner: Vadim Sobolevski</td><td colspan="4">Gross Risk</td></tr>
<tr><td>Risk Description</td><td>Likelihood</td><td>Impact</td><td>Total Score</td><td>Risk Status</td></tr>
</table>

15	Data submitted to FutureFlow is lost due to human error, resulting in the inability for FutureFlow to perform the processing.	2	3	6	Medium
	Existing Controls and Evidence	**Net Risk**			
	In on-premises deployments, Data Controllers (or Trusted Third Party) remain in charge of the physical infrastructure where the data resides and where the processing takes place. In these instances, Data Controllers remain in charge of data governance can apply industry-standard measures, such as access control, tiered privileges, and regular backups, to prevent data loss. Note: In planned Cloud deployments the FutureFlow platform will be part of a larger technical architecture and the processing will be subject to multi-institution governance. In such deployments, Data Lineage considerations will be delegated to the oversight bodies and the Cloud provider.	**Likelihood**	**Impact**	**Total Score**	**Risk Status**
		1	3	3	Low

Risk Ref.	**Owner:** Vadim Sobolevski	**Gross Risk**			
	Risk Description	**Likelihood**	**Impact**	**Total Score**	**Risk Status**
16	Data submitted to FutureFlow is damaged due to human error or hardware/software failures, resulting in the inability for FutureFlow to perform the processing.	2	3	6	Medium
	Existing Controls and Evidence	**Net Risk**			
	In on-premises deployments, Data Controllers (or Trusted Third Party) remain in charge of the physical infrastructure where the data resides and where the processing takes place. In these instances, Data Controllers remain in charge of the hardware and data governance and can exercise industry-standard measures, such as	**Likelihood**	**Impact**	**Total Score**	**Risk Status**
		1	3	3	Low

	regular backups and access control privileges, to mitigate the risk of data damage. Note: In planned Cloud deployments the FutureFlow platform will be part of a larger technical architecture and the processing will be subject to multi-institution governance. In such deployments, Data Security and Data Lineage considerations will be delegated to the oversight bodies and the Cloud provider.				

Annex A: Risk Tolerance Matrix

A heatmap/RAG Assessment which is used by FutureFlow when assessing risk which allows for an objective assessment of the overall risk. The level of overall risk will determine the treatment strategy used by FutureFlow/their clients. When using this matrix for assessing risk FutureFlow will replace the wording below with numerical values (e.g. Very Low=1, Low=2 etc).