

# Mobilising online data to fight financial crime

A guide to using AI in enhanced due diligence and investigations



Step 1 Orientation.4Step 2 Source and check official records5Step 3 News and media6Step 4 Wider internet data7Step 5 Social media9Step 6 Verification and onward research10Using AI to take due diligence to the next level11#XapienThem14How Xapien works15Using Xapien to fight financial crime16

### Introduction

## Your guide to using AI in enhanced due diligence and investigations

Online data plays an increasingly essential role in the fight against financial crime. The advent of the internet and the ensuing information explosion is a boon for analysts and investigators. It has transformed the transparency and traceability of sources of funds and fraudulent or corrupt activities.

At the same time, the volume of information can be overwhelming. At best this can lead to delays in customer onboarding, at worst, significant oversights.

When screening new clients or potential partners, fully effective due diligence searches deliver a complete and verified understanding of who a subject is. The UK financial services regulator, the FCA, lists 'using open source websites to gain a better understanding of the customer or beneficial owner, their reputation and their role in public life' as an effective EDD measure.

Its handbook goes on to note 'relying exclusively on commercially-available PEP databases and failure to make use of available open source information' as an example of poor practice in onboarding high risk customers.

But it can be hard to know where to start. A robust due diligence search should use the following sources:

- Official PEPs and sanctions lists and databases.
- The surface web, accessible via typical search engines like Google.
- The deep web, including corporate or organisation records.
- Unindexed online data. A huge amount of crucial information can be derived from sources such as grey literature and government reports which are not necessarily indexed by search engines.

Importantly it will require a well-trained analytical mindset to extract the right, relevant information from the right sources.

Here, we'll guide you through four fundamental steps for any truly effective enhanced due diligence, or background investigation process.

An effective due diligence process will go far beyond searching and collecting information. This information has to be read in detail to pick out nuggets of information and ensure that these nuggets link back to the original subject. Any due diligence search should be iterative, finding areas for onward research to build a fuller picture about the subject, and to corroborate existing unverified leads. A final, essential step is to distil and classify this information into a clear and concise report that can be shared amongst decision makers, and used for audit purposes.

A truly **intelligent** automated tool can be hugely helpful in increasingly the efficiency and effectiveness of this analyst work.





The advent of the internet and the ensuing information explosion is a boon for analysts and investigators.

## Step 1 Orientation

Before you start a deep investigation, it is often necessary to run some quick search engine-based searches on your subject. This enables you to scope and establishes some parameters for the investigation. It will allow you to build knowledge that will then typically inform all downstream strategies. It is also a great first step in building out you understanding for onward screening and verification of future data.

#### Strategies to help

This shouldn't take long. It is really about driving efficiency for onward, deeper dive research.

#### Use a search engine

At this stage a simple search on a well-indexed and common search engine is best. The objective is to get an overall sense of the scale and a better understanding of some key identifiers which will help in onward research and verification.

#### **Set limits**

At this stage it is worth setting clear limits and parameters as to the information you need to find, where, and by what time. These will need to be realistic to your organisation and objectives.

#### Store your data

To ensure simple tracking and referencing, store your findings in a central location and cross-reference them with other information you've gathered. Use a standardised reporting structure and timestamps on information to ensure a streamlined and audit-friendly process. An automated due diligence tool is helpful here.





## Step 2

### Source and check official records

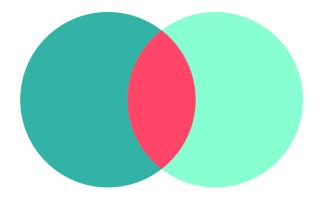
PEPs, sanctions and international watchlist checks are vital to assessing the regulatory and legal requirements associated with a subject. Although much of this data is freely available on government websites, each jurisdiction has its own list and manually checking each of these when onboarding is unsustainable. There are a number of vendors that offer automated screened solutions which check against consolidated databases.

Corporate and government records, as well as other so-called grey literature, can provide a rich account of a third-party's internal structure and governance. Depending on the jurisdiction, much of this data is available for public consumption, including financial filings and reporting, land registry records, and electoral roll data, and credit references. This enables verification of records that may have otherwise been obtained from the subject in question.

Such records can also provide critical insight into company and individual networks and sources of funds or wealth. Both of these essential sources of information can be disparate and challenging to collate, verify and analyse.



These essential sources of information can be disparate and challenging to collate, verify and analyse.



#### Strategies to help

Many jurisdictions require businesses and organisations to file various data and records which are then made available to the public. For example, in the UK, Companies House provides public access to business data, and in the USA, many business records are publicly available at federal level, state and local government level. You need to:

#### Assess what you can access

Assess what information is available in your region, what is related to your subject, and what needs to be accessed to complete a sufficient search. In most cases, the bulk of the information needed is publicly available. It's just a question of finding it.

#### Appeal for information

When you need information that is not publicly available, you can appeal for access. In certain jurisdictions, launching freedom of information requests can obtain information stored and processed by public sector bodies. In the United Kingdom this falls under the Freedom of Information Act and includes government departments, the armed forces, the police, the National Health Service, non-departmental public bodies, local government and educational institutions such as colleges, universities, and schools. This is only necessary in cases requiring deep investigations and your organisation will likely have a clear policy set out on extending such requests.

#### Seek local advice

Often official records require specialist expertise and relationships to find, interpret, and translate. Certain jurisdictions will require an in-person visit to a government office to pull records for example.

Whilst some jurisdictions do not host this data online, many do, and even where they don't automated tools can still help. Corporate data providers like Companies House, Open Corporates, Sayari, Dunn & Bradstreet, and Bureau van Dijk collate much of this information from disparate sources. Their structured databases are great resources for any researcher or investigator.

The challenge is consolidating and building on these sources and organising your findings into a simple, consistent and digestible report.

## Step 3 News and media

Vital insights can also come from wider, less structured, data, such as news and media articles.

When it comes to making an assessment of someone's reputation, and possible criminal or otherwise inappropriate associations and business practises, news stories provide an ideal starting point. Financial institutions and other organisations will often set their own standards, within the regulatory requirements, as to what an acceptable or unacceptable practice is. For example, a well-documented history of human rights abuses by an organisation may debar them from certain organisations' customer acceptance schemes.

Whilst traditional methods have focused on searching for negative coverage, there is also an opportunity here if the search is broadened. Media coverage of all complexions can add real value to your understanding of who you're dealing with.

News stories can also contain a wealth of information which will point you in the direction for further investigation or onward research. For example, when conducting enhanced due diligence on a British resident, you naturally check Companies House for the aforementioned corporate records. However, through news screening you may find that they were connected to a large property purchase in Mali, a country listed by FATF as having 'strategic deficiencies' in its AML controls and under sanctions in some jurisdictions. This will require further investigation and verification of the circumstances surrounding the alleged purchase prior to onboarding.

It is often necessary to expand screening and searches to foreign-language sources. The objective here is to gain a broad contextual understanding of the third party by analysing their present and past actions and associations, as well as their public image.



Media coverage of all complexions can add real value to your understanding of who you're dealing with.

#### Strategies to help

Google and other search engines are often the first go-to methods for researching published media.

#### Filtering tools

Google and other search engines incorporate a number of built-in filters that allow users to search by date, language, and region. This can be hugely valuable when looking for media and risks associated with international partnerships.

#### Use a variety of search engines

Google is great. But its commercial search engine results page (SERP) features such as Ads, Knowledge Graph, Snippets and Map Pack can crowd out relevant results, making it tricky to find the information you need. It might also not the best platform when dealing with international subjects. Other search engines that you need to know about include:

- Yandex is best for Russia and CIS coverage.
- Baidu is best for China coverage.
- **Duckduckgo** is best for privacy and a clean perspective.

Changing your own location and language settings within search engines can also be valuable here. For example, if I set my location to France, and language to French on Google and run a news search then French media results will rank higher.

#### Paywalled media

Unfortunately using free search engines cannot be the sole strategy in this step. A large number of influential media outlets operate on a subscription basis and are gated behind paywalls. Your organisation will likely have a policy, and subscriber accounts with relevant ones.

#### News and media databases

When Google is not enough, there are plenty of premium news and media databases optimised for research purposes. These include **LexisNexis**, **Factiva** and **NewsEdge**. These have built-in advanced filtering systems, and subscription options can provide access to otherwise paywalled media.

## Step 4 Wider internet data

The buck doesn't stop with media searches. The internet contains a vast amount of other information about an individual or company that could be relevant to your search. The challenge is finding, reading, verifying, and making sense of it all.

Beyond the marketing material, company websites can provide crucial insights into an organisation or an individual associated with them, with their legal statements, privacy notices, and policy statements sometimes buried deep on the website.

Blog posts might discuss controversies, accusations against or hearsay about your subject that flag otherwise undetected risks to you. Blogs and online forums could contain rumours of wrongdoing of conspiracy theories about an individual or organisation that has not yet been reported on by news sources, but could be an area of potential risk in future. Make sure to assess the credibility of your sources and cross reference information from a blog post or forum against other, more credible, outlets. However, by considering all available online sources, you ensure you are getting the whole picture about your subject, and assessing potential future risks as well as current risks.

Other sources include databases provided by subscriptions services such as **RocketReach** which stores email addresses and phone numbers. Platforms like **Pipl** provide an index of identity information that cross-references various internet sources, listings, directories and archives for identity verification.



Beyond the marketing material, company websites can provide crucial insights into associated organisations or individuals.



#### Strategies to help

Search engines are a key entry point here. But remember that they will only display indexed pages, not all internet data. This will require specialist tools.

Even on search engines, it is usually necessary to employ advanced search techniques to extract deep insights that go beyond the first few pages of results. Remember, the negative news story, or critical investigative lead obtained from an obscure online post could be decades old. Consider options such as:

#### **Targeting functions**

Counterintuitively, often the fewer results you get on Google, the better, as this means you have inputted the specific search terms relevant for your subject. Millions of results clutter the space. Ten specific results that relate to your subject are gold dust.

At the very least use quotation marks in your searches. Searching for 'John Smith' will dramatically reduce the number of search results you have to trawl through and will only show results with that specific name in that order. Without quotation marks you'll end up with all the results for 'John' and all the results for 'Smith'!

Making use of Boolean operators – are, and, or, not or +, -, () – also helps structure searches for more productive results.

#### **Specify sources**

The **filetype:** function will return results for your specified filetype plus the search term or site you list. For example, the search **filetype:pdf 'Official List'** can help you find official lists and records for verification purposes.

The **site:** operator will find only information from a specific site. This can be really helpful for sites that don't include their own search function. For example, if I want to find out about risk on Xapien's site, I can enter **site:www.xapien.com 'risk'** and jump straight to the pages I'm interested in.

#### Advanced search techniques

Search engines often hold capabilities for advanced searches that can be triggered through using specific terms.

For example, adding the code &tbs=qdr:m to your results URL can deliver you filtered results from the past month. Slight amendments can be made to the code to change this time period as and when needed.

Wildcard character operators (\* or **Shift + 8**) substitute an unknown variable for your search. It can be used if you don't know a subject's first, middle, or last name but have other information to help you search for them.

#### Additional resources

Check out Google's guide to refining web searches.

Investigative organisation, **Bellingcat**, have a really handy online toolkit.

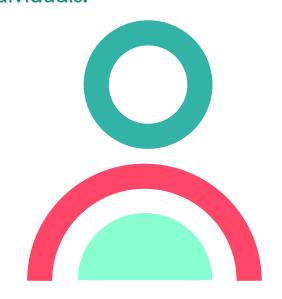
Michael Bazzell's <u>Open Source Intelligence</u> <u>Techniques</u> is considered the Bible of online research (or visit his website).

The right solutions sit above these sources and incorporate other tools enabling you to automate much of the time-consuming work associated with the research process. Purpose-built due diligence tools such as **Xapien** conduct a wide online search, including anything a search engine or database search would produce.

Xapien integrates other resources and tools and brings them into an effective due diligence structure. Xapien will not only do the grunt work of gathering vast amounts of data, but it will read that data, disambiguate, filter, and analyse it for you. Just like a human, it incorporates techniques from the Michael Bazzell, Bellingcat and other schools of online research. You don't need to worry about what might have been on page 300 or some other obscure source. If it includes relevant insight on your subject, Xapien will report it.



Company websites can provide crucial insights into associated organisations or individuals.



### Step 5 Social media

Social media is an incredibly large and dynamic source of information that provides a rich, timesensitive account of individuals, events, and associated entities. Neglecting its usefulness in modern due diligence is a major oversight. Not only does social media help analysts build a perception of a third-party's networks, it also provides insight into sentiments towards them.

That said, every business and organisation will have its own policy on how to approach researching social media data. How far this research has to go depends on factors like risk tolerance, privacy concerns, and what information is uncovered that may require further investigation.

The huge volume of social media can be useful, but information can get lost in the mix. False positives are a persistent issue and information from social media cannot be trusted without further verification and supporting evidence. In order to gather and cross reference vast quantities of social media data effectively, advanced investigative tools like Videris and Maltego are necessary.

Technology platforms with **verification** and **disambiguation** capabilities can help to spot inconsistencies and discrepancies to enhance accuracy.



"

Not only does social media help analysts build a perception of a third-party's networks, it also provides insight into sentiments towards them.

#### Strategies to help

Whilst social media investigations should respect boundaries of privacy, the volume of social media data published for public consumption is vast. You can use this information to:

#### **Draw connections**

The wide variety of social media platforms available means that most people will have multiple accounts. Get a bigger picture of your subject by drawing connections between those accounts and who they are connected with.

Free tools like Namechk.com and checkusernames. com can help you find new accounts under the same username. For example, if you know that Joe Bloggs uses @JBloggs2 as his Twitter handle, he may have other accounts using the same prefix. This can help you to broaden your search, and get a sense of old blog/social accounts that may expose relevant pieces of information about the person you are looking into.

#### Check timelines/locations

Tracking social media data can provide information on the history of a subject's actions and endeavours, including when and where they may have been previously based. When conducting source of wealth or source of funds investigations, photos and other posts on social media may indicate when a particular asset was acquired.

#### Assess online reputations

Similarly to conducting negative news screenings, social media provides a wealth of insight on a subject's reputation. Posts on online platforms give an idea of how an individual/organisation likes to present themselves to the wider world and, in turn, how the wider world perceives them.

## Step 6 Verification and onward research

Verifying the data you have collected from the steps outlined above is the final step in any comprehensive due diligence or investigative process.

Verification is typically painstaking and requires a high volume of manual work. Uncovering connections whilst avoiding false positives is tricky. The time lost in checking that level of detail can cause delays with onboarding new customers and partners.

Nevertheless, performing due diligence searches without a rigorous strategy for verification not only slows down customer acceptance, but creates a hotbed for future litigation and other risks.

The verification process also enables the analyst to build knowledge about their subject. This is part of the overall process of effective due diligence, which should be iterative. Whenever a new piece of information is found about a subject it should be fed back in to the beginning of the research function to see if more information can be found off the back of that. Each of these pieces then need further verification, and onward research, and so it goes on.

Manually processing potentially vast quantities of documents and data is time-consuming and the enduring risk of human error threatens to undermine the due diligence process. To offset that, purposebuilt due diligence and background reporting tools assist researchers in finding the connections between disparate sources with the guarantee that those connections have been verified.



Verification is typically painstaking and requires a high volume of manual work.

#### Strategies to help

Every step in the research process must ensure that the information collected is reliable and accurate. The key to guaranteeing effective verification and onward research is through gathering a broad range of data and keeping track of the sources of that data.

The challenge comes in ensuring that the final report remains digestible and user-friendly. Concise findings can then be used for onward decisions after due diligence. This can be done via:

#### An iterative research process

The due diligence process itself should be sequential and cyclical in its procedures and outcomes. Any analyst or tool must take single pieces of information, verify them and build on them. It should feed new, verified information back into the process in order to build a full picture of the subject.

#### Simple data visualisation

Reporting that information has to be as, if not more, efficient than gathering it. This in itself can consume a vast amount of analyst time. The use of a template that takes into account the multiple categories of relevant data will make for an exhaustive yet comprehensible report.

#### Multi-access capabilities

The final reports need to be shareable and accessible for as many people as necessary to the due diligence process. This also provides all team and organisation members full risk awareness when making onward decisions. Templates, as highlighted above, play a critical role in this. New cloud-based solutions can transform the speed and ease at which information can be shared.

#### Full provenance

Confidence in the information uncovered depends on the sourcing. Allegations made in a blog or social media post have to meet very different evidential standards to something reported in a leading national paper, or an official document. The number of sources reporting an allegation is also significant. Reporting must include details of the provenance of information in order to justify onward decisions. Detailed sourcing is even more important when it comes to using automated tools. The human researcher will need assurance that the tool has found reliable information.

## Using AI to take due diligence to the next level

Human teams have their work cut out if they want to produce genuinely comprehensive background reports fit for modern due diligence. That gap is precisely what automated and AI-assisted platforms can fill.

Automation can streamline the process of working with vast volumes of data, enabling analysts and researchers to focus their energies where it matters. Automated search technology and supportive AI tools for processing data transform the modern due diligence process, replicating the full human research process. This allows human teams to focus on high-level tasks that drive additional value and productivity. The outcome is clear, detail-rich accounts of prospective partners that enable faster and more informed decisions.

High-quality due diligence processes provide value beyond identifying risk. They help build a strong understanding of the third party which can assist in building a solid relationship with them in the future.

There are several benefits to employing automated technologies in the due diligence process:

#### 01 Standardised due diligence

Standardising due diligence processes is a matter of both efficiency and certainty. It puts a system on an otherwise freeform process that is dependent on different analysts and their respective research training, tenacity, and capacity. Utilising a singular platform helps businesses and organisations develop long-lasting protocols and standards for their due diligence.

#### Advanced record-keeping

Manual methods are inconsistent and often rely on one or two individuals keeping track of their research. This data might be recorded in an insecure environment. Automated due diligence platforms are self-contained and offer a singular ecosystem for the entire research process, input, and output.

#### **Ensure consistency**

No two analysts go through the exact same process, no matter their training. This becomes more challenging the bigger your organisation is. Using machine-reproducible techniques and cloud computing scale enables repeatability and consistency across even the biggest teams and processes.

#### Avoid human error

The speed and highly-informed insights achieved by due diligence platforms don't come at the sacrifice of detail. It's a matter of scale and AI is exceptionally good at working with volumes of data beyond the capacity of human teams. Simply put, AI doesn't get distracted, or burnt out. It will keep searching until it finds the information you need.





Automated search technology and supportive AI tools for processing data transform the modern due diligence process.

#### 02 Efficient due diligence

Manual due diligence is notoriously time-consuming and inefficient, even without the assumption that the final data is up to standards.

#### Maximise resources

Automated due diligence removes many of the menial tasks associated with manual searches and data processing. It makes all the small decisions so that humans can make the big ones. The result is a faster, more efficient due diligence process that works smarter and harder whilst maintaining a critical level of insight. Automation does not replace human efforts, but enables them to use their skills and energy where they are most needed, maximising productivity.

#### Gain self-sufficiency

Automated due diligence allows organisations to expand their in-house capacity by optimising their analyst teams. The speed, simplicity and standardisation that automation brings can also be critical when conducting periodic research or audits, annual reviews, etc.

#### 03 Comprehensive due diligence

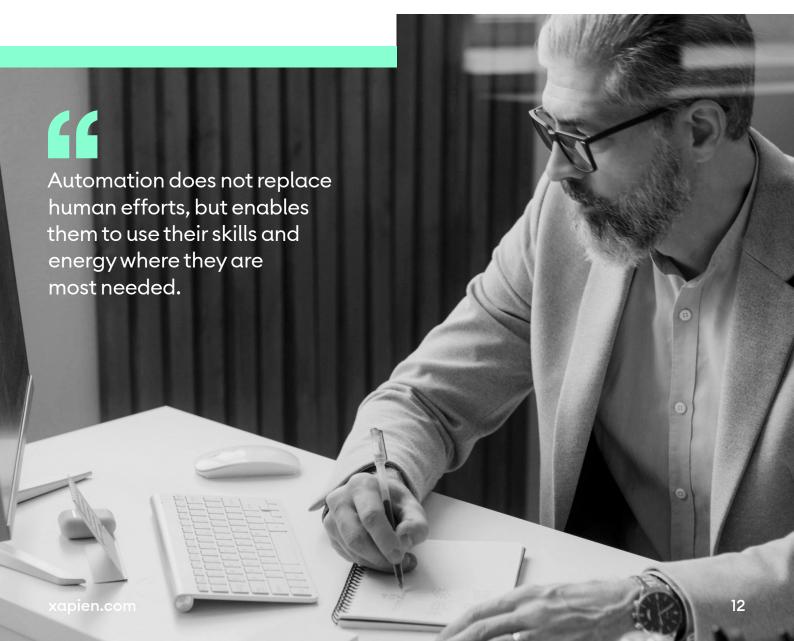
Speed and efficiency should not come at the expense of detail. Automated due diligence also returns a far richer account than is practical via manual techniques, when the quantity of information to check is so vast.

#### **Guarantees detailed insights**

Automated processes iterate, cast a wide net, gradually narrow the search, and filter out false positives. This enables a complete, nuanced understanding of a potential client, customer, hire, investor or donor. This process, though possible, is impractical for most organisations to do on a regular basis via manual searching.

#### Builds a better understanding

Information discovered from the truly comprehensive report that automation provides is useful beyond the due diligence process. It can assist in building a strong relationship with the third party.



#### 04 Extended due diligence

Whilst typical surface web sources such as media outlets are readily available through Google and other search engines, the process of search engine optimisation (SEO) biases results towards websites with higher outreach. This also neglects the vast quantity of other, unindexed data, or data indexed on other, obscure search engines.

Internet users produce roughly 2.5 quintillion bytes of data a day. 79 zettabytes of data alone was consumed globally in 2021 – that's about 1,200,000,000 times the size of the Google Earth database. And that number is expected to increase to over 180 zettabytes by 2025.

An automated due diligence platform takes due diligence forward by targeting all sources as opposed to a limited few. The benefits of this wider approach include:

#### Knowledge

Using automation, it is possible to crawl through huge volumes of information from far beyond Google's top search results, stitching together sources from multiple sources, regions and languages. This offsets any gaps in information that may have resulted from a lack of access to the crucial data needed. It enables verification which enables decisions to be made on a factual, knowledgeable basis.

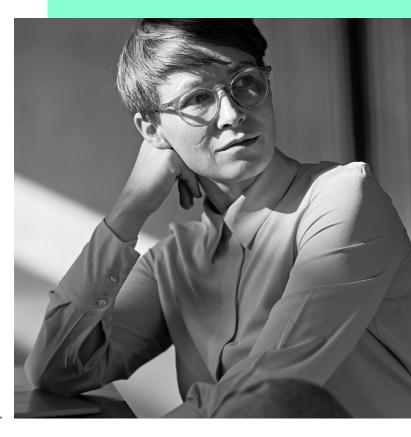
#### A full and fair picture

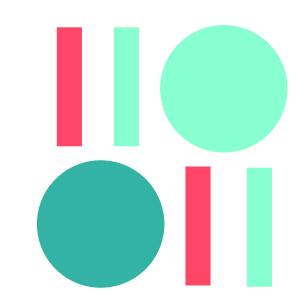
Beyond the surface web lies a vast volume of nonindexed data hosted on deep web databases and networks. Whilst much of this data is publicly available and intended for public consumption, it is near-impossible to locate using standard search tools.

The future generation of truly intelligent due diligence solutions use advanced natural language processing techniques to read 'structured' and 'unstructured' data. Simply put, existing tools rely on lists and categorised information such as corporate records. Next generation solutions combine this with information gathered from the media, blogs, social media and any number of other information sources that do not clearly categorise insights. This enables you to get a genuinely full and fair picture of the individual or organisation you are assessing.



Using automation, it is possible to crawl through huge volumes of information.





## **#XapienThem**

Xapien has developed a cutting-edge tool to meet the demands of modern due diligence.

Our automated platform supercharges due diligence by streamlining the process without sacrificing the details. In fact, Xapien advances the due diligence process by extending searches beyond known lists and databases into any form of structured and unstructured internet data. This includes multilingual sources, corporate data, the deep web, blogs and articles and everything in between. Xapien provides organisations with reports that are broader and deeper than possible via traditional methods, faster.

Launching due diligence searches with Xapien is as easy as inputting the name of your subject, adding some context, and hitting go. The platform then crawls through all online data, using powerful Al algorithms to identify the information that is relevant to you and your subject.

The results are displayed in a single, shareable report alongside their original sources. Xapien automatically identifies assets, associates, wealth, business roles, descriptions, quotes and affiliations whilst providing avenues for further exploration.





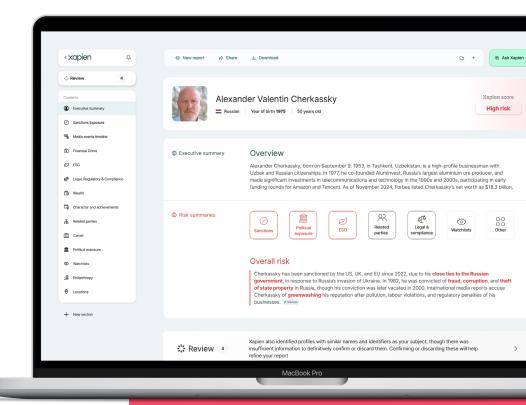
## Seeing the product was a Eureka moment.

**Head of Claims Fraud** Zurich Insurance



### I cannot overstate how much quicker due diligence is with Xapien.

MLRO Griffin Financial Technology



## How Xapien works

#### **Cutting-edge AI platform**

Xapien is an advanced automated AI platform built and optimised for due diligence and background research. It is not a single tool, but a suite of tools that harness powerful machine learning algorithms to search, categorise, and analyse data at scale.

#### Search and collect

A Xapien report is the product of thousands of searches run simultaneously in many different languages. Xapien is greedy, collecting every relevant bit of data it can find online to compile a huge knowledge graph. While this search and collection process only takes our system minutes, it would take a human analyst hours, days, and even weeks to search across all of these sources.

#### Knowledge extraction

Rather than extracting raw data, Xapien extracts knowledge. Our natural language processing and machine learning algorithm, is capable of making human-like interpretations of text at massive scale. The result is a comprehensive, fact-based picture of your subject, their wider connections, associations and reputation.

#### Verification and disambiguation

Xapien features a powerful disambiguation engine that resolves the identities of subjects using probabilistic algorithms. The information that Xapien discovers is modelled as 'possibly true' until its algorithms can positively verify the likelihood of identities and connections. This means fewer false positives and faster onboarding.

#### Risk analysis

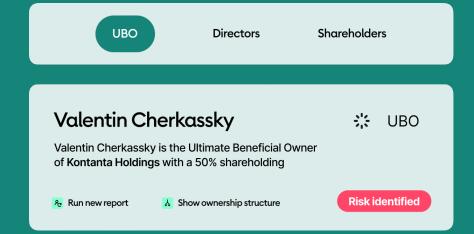
Xapien goes beyond searching and categorising information. Our Neurisk algorithms understanding the natural semantic usage of words and terminology like a human would. This means that more abstract inferences of risk are not ignored. They identify each mention of your subject and work to classify the tone and sentiment of information associated with it. Analysts can quickly survey reports and see where mentions of the subject involve any connotation or inference of risk. This catches more risks, streamlines onward investigations, and enables genuinely risk-based final decisions.



Together with Xapien we are pioneering the future of risk mitigation.

EVP and General Manager, Risk and Research Dow Jones







Xapien effectively filters out false positives and provides high-quality information.

General Counsel
Pinsent Masons

## Using Xapien to fight financial crime

Founded on the principle that everyone deserves to know who they are doing business with, Xapien democratises access to valuable insight. For too long the battle against financial crime has been effectively a game of catch up. By providing genuine insight into the people you are doing business with, Xapien can be used to identify and shut down criminal networks and fraudsters.

Its use of structured and unstructured data is invaluable. Traditional screening and watchlist checks are no longer sufficient. But effectively mobilising the vast and every-growing expanse of online data is humanly impracticable by manual searching alone. This is where Xapien steps in.

As a cloud-based solution that requires no training to use or understand, Xapien can be rapidly integrated into any organisation.

Join the fight against financial crime, avoid regulatory and reputational risks, streamline your onboarding, forge fruitful relationships with sustainable and strategic clients and most importantly, foster a culture of trust and authenticity in your organisation. #XapienThem.

#### Get in touch

Discover how Xapien can transform your use of online data today.



By providing genuine insight into the people you are doing business with, Xapien can be used to identify and shut down criminal networks and fraudsters.



