COMMUNICATIONS INTELLIGENCE:

Creating the Foundation for Your Organization's Communications Data Strategy

By Brandon Carl
Executive Vice President, Product Strategy
Smarsh

Table of Contents

Communications Intelligence

Introduction

The Communications Revolution

From cave paintings to palm pilots, and rotary phones to Reddit threads, the format of our communication has changed significantly. But the need to connect with fellow human beings persists. It is one of the most powerful tools we have.

How people communicate continues to evolve. Innovations from the last 20 years alone may be more consequential to human interaction than perhaps the entire century before it. In the business world, high-speed internet, search engines, and collaboration tools have fundamentally altered how many companies operate. Businesses can expand globally, support remote work, and build a brand — entirely online.

As a result, there has been an explosion in both the volume and variety of communications that are generated every moment of every day. Water-cooler chats and conference room discussions became Slack chats and Zoom calls. What was previously a passing hallway conversation is now stored in the permanent record of corporate communications.

Yesterday's Approaches Don't Solve Today's Problems

These changes have profound implications for corporate responsibility, legal and discovery risks, and the operational viability of compliance programs. This is particularly true in the financial services industry, where FINRA-regulated and SEC-regulated organizations are required to preserve, monitor and manage business communications.

For several years this could be accomplished through email journaling or transferring communications to an on-premise archive. But most archiving tools did not plan ahead for new social media platforms or chat applications that would one day be adopted for business communications.

Over time, many organizations have found themselves with a jumble of disparate tools and data in multiple formats — making it difficult to efficiently and effectively manage business critical compliance. The historical archive may be capable of accommodating books and records requirements. However, it is ill-suited to meet the scope of today’s communications generated among employees and with customers, across platforms, devices and geographies, around the clock.

Communications Intelligence: The Solution For Tomorrow

Fortunately, breakthroughs in natural language processing — coupled with advances in cloud architecture — provide the foundation needed to rise to the challenge. What's more, communications are a strategic asset for any company, pointing to risks and opportunities you would never have been able to see without AI-enabled technology. Company communications offer an often-underutilized lens for viewing business value, cultural shifts and red flags — warranting more attention than just regulatory obligation.

Successful financial organizations must rethink and redesign core processes for communication governance and risk management in today's business environment. What will set them apart is developing a comprehensive **communications data strategy** to meet evolving regulatory requirements and uncover the value within the communications they are required to retain.

To implement and enforce a communications data strategy, firms' technology needs to evolve from the compliance archive of yesterday to a solution that is truly appropriate for today's communications, technology and business landscape. That's why we have introduced a new market category: Communications Intelligence.

In the sections that follow, we outline:

- Why a communications data strategy is critical for modern enterprises
- How Communications Intelligence expands beyond traditional information archiving
- How to lay the technology foundation for the future of your enterprise's communications data strategy

Defining Communications Intelligence

The average Fortune 500 company sends and receives millions of communications per day. These communications contain multitudes of signals, both independently and in aggregate: untapped revenue opportunities, costly operational errors, security threats, real-time cultural indicators, and compliance and brand risks. Unfortunately, most companies have no systematic way to surface these signals.

Communications Intelligence is built on 3 key principles:

1. Communications are a strategic asset that must be retained and managed for governance purposes
2. Communications data differs from transactional data and requires a unique set of technologies to be managed and governed appropriately
3. Only machine learning can surface critical signals at the speed and scale that your business requires

Communications Intelligence enables you to capture, retain, analyze and act on the signals in your communications that are most critical to your business. Specifically, it comprises the strategies and technologies used by enterprises for the collection and analysis of human communications data to help quickly identify risks, recognize new business opportunities, and improve operational strategies. This infrastructure is the technological foundation for use cases such as e-discovery, internal investigations, communications surveillance and more.

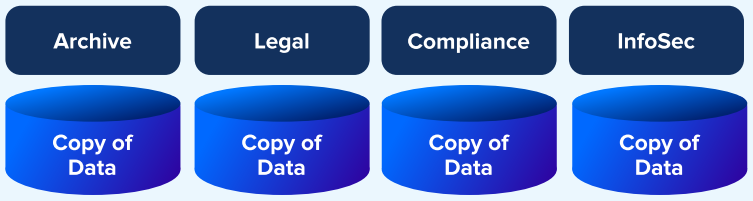
Why is Communications Intelligence important?

A well-designed Communications Intelligence infrastructure enables firms to:

1. Uncover hidden risks and opportunities
2. Cost-effectively scale to the speed and volume of today's communications
3. Safely govern sensitive data

Before

- **Duplicative Costs**
- **Higher Security Risks**
- **Siloed Insights**



After

- **Shared Costs**
- **Lower Security Risks**
- **Centralized Insights**

1 Uncover hidden risks and opportunities

Despite strong corporate cultures and extensive training, people still make mistakes. Left undiscovered, bad actors within companies will continue to mistreat or harass their coworkers, the firm and its customers. Even well-meaning people misbehave; studies show that up to 59% of departing employees steal company data after they were laid off, fired or quit their jobs.[1]

Communications data contains predictive and postmortem evidence of these risks that go far beyond what metadata systems can do. Unfortunately, most companies have no systematic way to surface and act on these business-critical insights. Some real-world examples:

Financial crime

A bank executive was charged by the SEC for insider trading, as evidenced through messages the executive sent to a friend via the encrypted WhatsApp messaging platform.[2]

E-discovery

In an intellectual property dispute alleging the theft of trade secrets, Calendar Research v. StubHub, a judge issued the production of relevant Slack messages from the defendants.[3]

Intellectual property theft

Waymo discovered that an employee, Anthony Levandowski, had stolen more than 14,000 design files and sold trade secrets to Uber when it received an accidental email from a supplier.[4]

Sexual harassment

Following reports of "rampant" sexual misconduct, harassment and discrimination throughout the workplace, video game company Activision Blizzard is being investigated by the SEC and sued by a California civil rights agency.[5]

Remote work

The Financial Conduct Authority (FCA) in the UK released a list of "Remote or hybrid working expectations for firms," which outlines how financial firms can assure adherence to regulatory obligations with a remote or hybrid working staff. This includes the ability to maintain recordkeeping procedures, review call recordings, and perform order and trade surveillance.[6]

Many firms are looking beyond traditional risks and using communications to uncover hidden opportunities.

FOCUS: IDENTIFYING AND ACTING UPON RISK

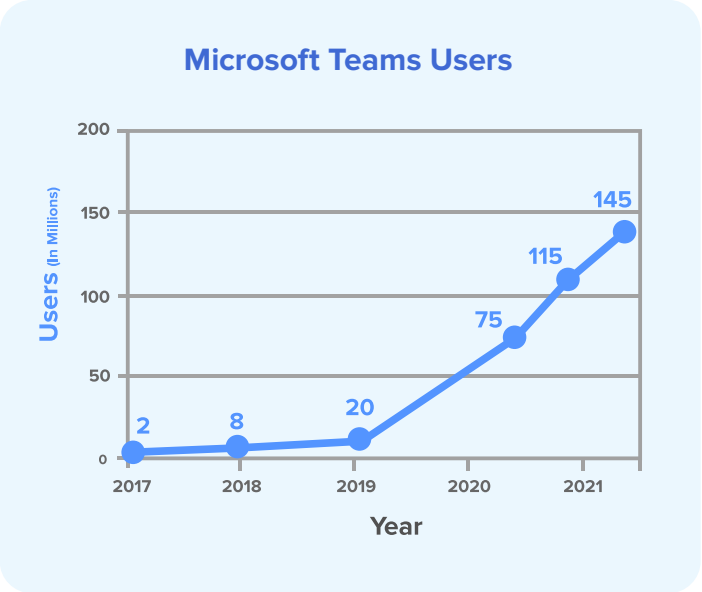
Organizations need to capture and preserve information in order to manage risk in many forms:

1) https://docs.broadcom.com/doc/ca-data-content-discovery-infographic
2) https://www.smarsh.com/blog/thought-leadership/silicon-valley-banker-charged-with-insider-trading-used-whatsapp-messaging/
3) https://casetext.com/case/calendar-research-llc-v-stubhub-inc-1
4) https://www.theverge.com/2017/2/23/14719906/google-waymo-uber-self-driving-lawsuit-stolen-technology
5) https://www.theguardian.com/technology/2021/sep/21/activision-blizzard-confirms-sec-investigation
6) https://www.fca.org.uk/firms/remote-hybrid-working-expectations

2 Cost-effectively scale to the speed and volume of today's communications

"Work from anywhere" has created an explosion in speed, scale and types of business communications. Exponential growth in tools such as Microsoft Teams and Zoom have changed how we communicate from email-centric to truly multi-modal.

Over time — and with the steady uptick of data — more communications-based use cases have arisen across departments. Security and HR teams conduct internal investigations. Client-facing teams may mine the data for customer insights. These efforts are often siloed. At best, these stakeholders are gluing together a complex web of data types from multiple vendors. At worst, they are paying for overlapping infrastructure and creating data privacy and security concerns.



Microsoft Teams daily active users rose from 2 million in 2017 to 145 million in 2021[7]

Traditional methods of oversight cannot handle the scale of modern communications. It is no longer viable to use basic search terms in e-discovery, keyword spotting in customer calls, and random sampling in compliance checks. These approaches are costly and labor intensive, result in a prohibitive false-positive count, and fail to surface the most critical insights.

Machine learning and multilingual natural language processing can solve the scale issue. Owing to breakthroughs in both fields, modern technologies can "understand the semantic meaning" of language and elastically scale to nearly any workload. This enables cost-effective, timely responses and risk management.

7) https://www.businessofapps.com/data/microsoft-teams-statistics/

3 Safely govern sensitive data

Your communications contain some of your firm's most sensitive information. Each time you export this data to another application (e-discovery, surveillance, etc.) you are creating potential breakdowns in your ability to safely govern the data.

Furthermore, GDPR and the California Consumer Privacy Act define the "right to be forgotten." Should a demand arise to delete data, **every** application in your ecosystem that consumes communications data must abide by these rules and regulations.

By adhering to best practices in your technology stack, you can mitigate privacy and security concerns while quickly maximizing your time to value. This includes thoughtful information governance, security practices and privacy adherence.

RECOMMENDED READING:

Managing Global Data Privacy Laws and Communication Regulations in Financial Services

How does Communications Intelligence differ from information archiving?

Communications Intelligence is an evolution of Enterprise Information Archiving. Historically, data was stored primarily to meet search-ready archiving and supervision processes (“books and records”) required for compliance, and to have on hand for legal purposes.

Traditional information archives can no longer meet the scale, speed and expanded stakeholder needs modern-day enterprises require. Executives should re-orient their thinking towards a comprehensive communications data strategy to power the future of their business. Communications Intelligence meets regulatory requirements via WORM-compliant data stores. More importantly, it expands the power of your communications data.

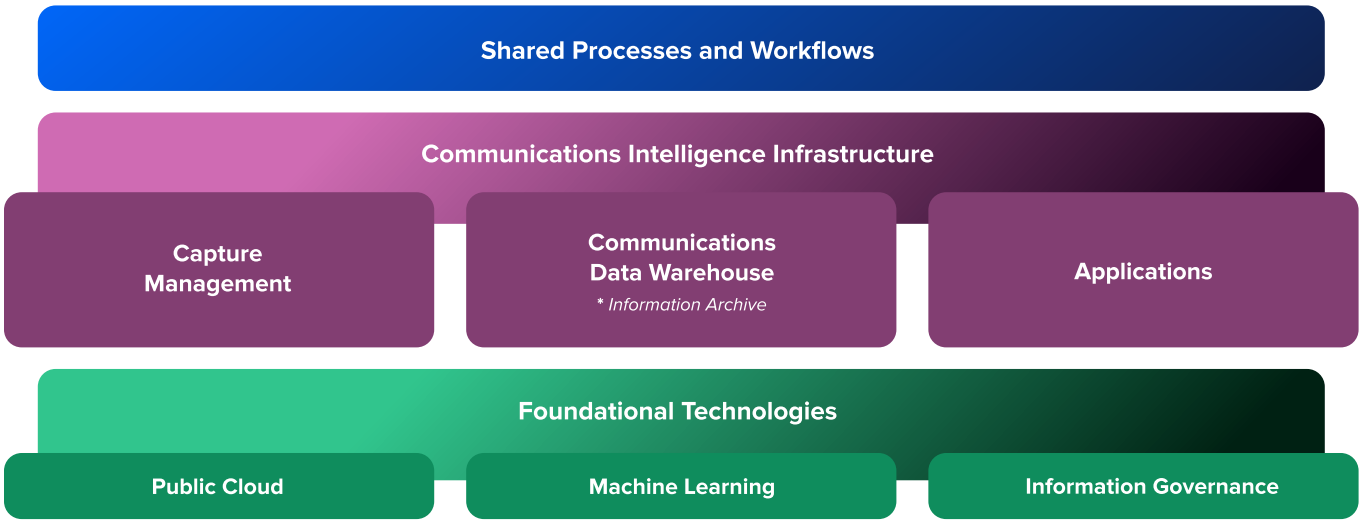
The table below outlines the key differences between traditional Information Archiving and Communications Intelligence.

From Information Archiving	To Communications Intelligence
Information Archive	Communications Data Warehouse
Closed, Export-Driven Data Architecture	Dynamic API-Driven Data Access
Books + Records Only	**Books, Records + Analytic Enrichment**
Flattened Email-Centric E-comms	Full Fidelity Email, Chat, Social, and more
Separate E-comms and Voice Archives	**Joint E-comms, Voice, Video Archives**
Keyword Search	Keyword and Semantic Search
Limited Machine Learning	Deep Learning and GPU-Driven ML
No Model Governance	Systematic Governance + Drift Detection
Siloed ML Insights	Shared ML Insights
Private Server Infrastructure	**Scalable, Public Cloud Infrastructure**
Active-Passive w/ BCDR	Always Available Georedundancy
Disconnected Communications	Connected Communications Graph
Privacy as an Afterthought	Privacy by Design
Batch-Driven t+1 Processing	Streaming Insights
Company-Specific Machine Learning Models	Industry-Coordinated Machine Learning Models
Compliance and Legal Stakeholders	**Expanded Set of Stakeholders**

How does it work? The Communications Intelligence technology stack

Communications Intelligence seeks to securely process all relevant communications, in key languages, at your company's scale. To maximize business value, this requires tight integration of three top-level components: communications capture, information archiving and end-user applications that work with communications data. This must be built on a firm technology foundation rooted in industry best practices. It must also provide common workflows to enable coordination across business units.

We discuss each of these below and conclude this section with a reference architecture.

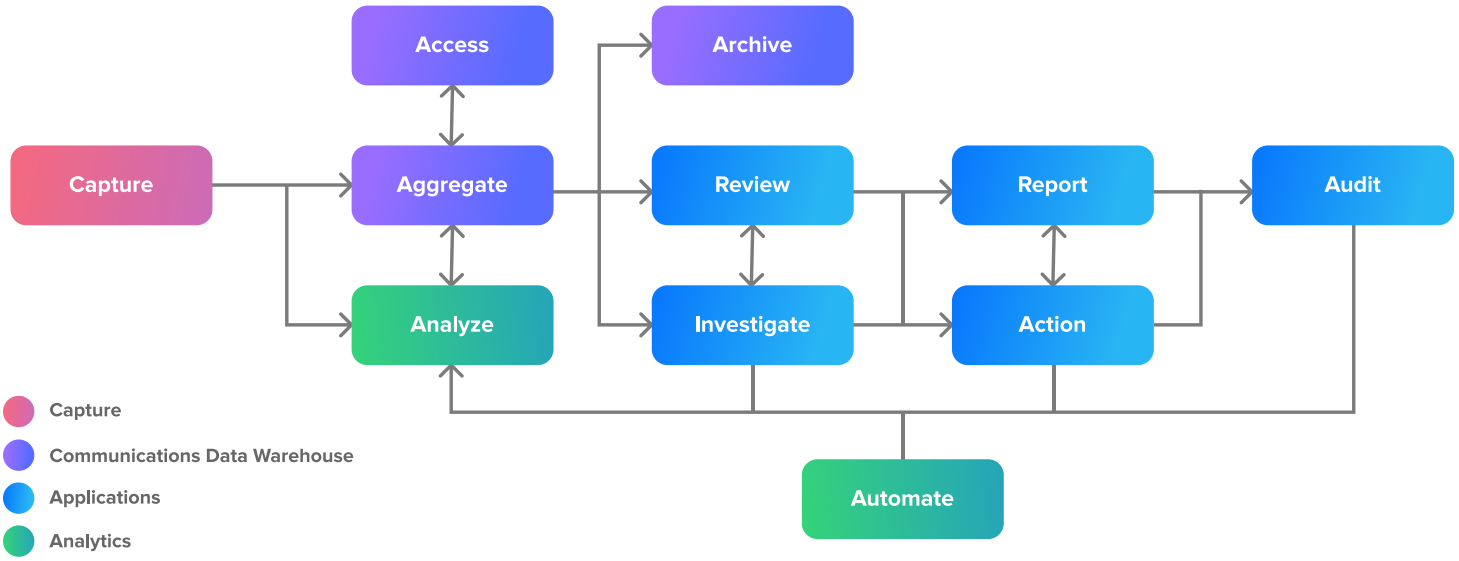


Foundational Technologies	
Scalable cloud infrastructure	The underlying infrastructure must be able to cost-effectively process, analyze, store and retrieve a company's communications. For large organizations, public cloud infrastructure will be the best option to ensure future success. An effective infrastructure provides: • Deployment zone coverage in relevant geographies • Global network infrastructure • Elastic on-demand computing • Multi-petabyte storage scale • Data redundancy • High availability
Machine learning focused on human communications	In the past few years machine learning has established itself as the only reasonable and effective way to process communications data. Rule-based methods generate too much noise and human staff are inconsistent and cost ineffective. Effective machine learning capabilities include: • Elastic compute (CPUs and GPUs) • Human-in-the-loop active/guided learning for machine learning model training • Bursty machine learning model backtesting • Natural language processing (NLP), including transcription and translation • Unsupervised learning capabilities such as clustering and language modeling

Information Governance	As a strategic asset of your firm, it's important that your communications data is secure and that privacy regulations are appropriately respected. Appropriate privacy, security and compliance provides: • Physical security • Access management • Double encryption at rest and in transit • WORM compliance • Hardware refresh immunity • Compliance with GDPR, CCPA, and other data privacy regulations • Customer-managed encryption keys
Communications Intelligence Infrastructure	
Capture Center	The Capture Center connects to your communications applications to ensure that your data is captured, routed, standardized and reconciled across all your key communications channels.
Communications Data Warehouse	After your communications data is captured, it is indexed, analyzed and retained in the Communications Data Warehouse, which serves as the center of your communications ecosystem. While a Communications Data Warehouse can meet the regulatory requirements of a compliance archive (such as WORM compliance), it also enables higher frequency, bilateral access patterns to derive, access and store critical insights.
Intelligent Applications	Applications act on the communications data and corresponding insights to help your company. Most companies start with communications surveillance or e-discovery and extend to powering a broader set of use cases. Across these components is an Analytics Engine, which detects and surfaces key signals within the communications data. Most commonly this will leverage a combination of rules, filters and natural language processing applied to individual communications as well as the broader communications graph.
Shared Processes and Workflows	
Review Workflows	Message-centric views of risk. Enables firms to identify, prioritize and review relevant communications.
Insights and Investigations Workflows	Person-centric views of risk. Enables firms to traverse a timeline and network of communications and to bundle relevant signals into cases.
Reporting and Audit Workflows	Information summarizing program effectiveness and efficiency. Enables firms to understand, manage and audit their business functions.

A deeper look into the components outlined above shows that most applications of communications data follow a similar workflow:

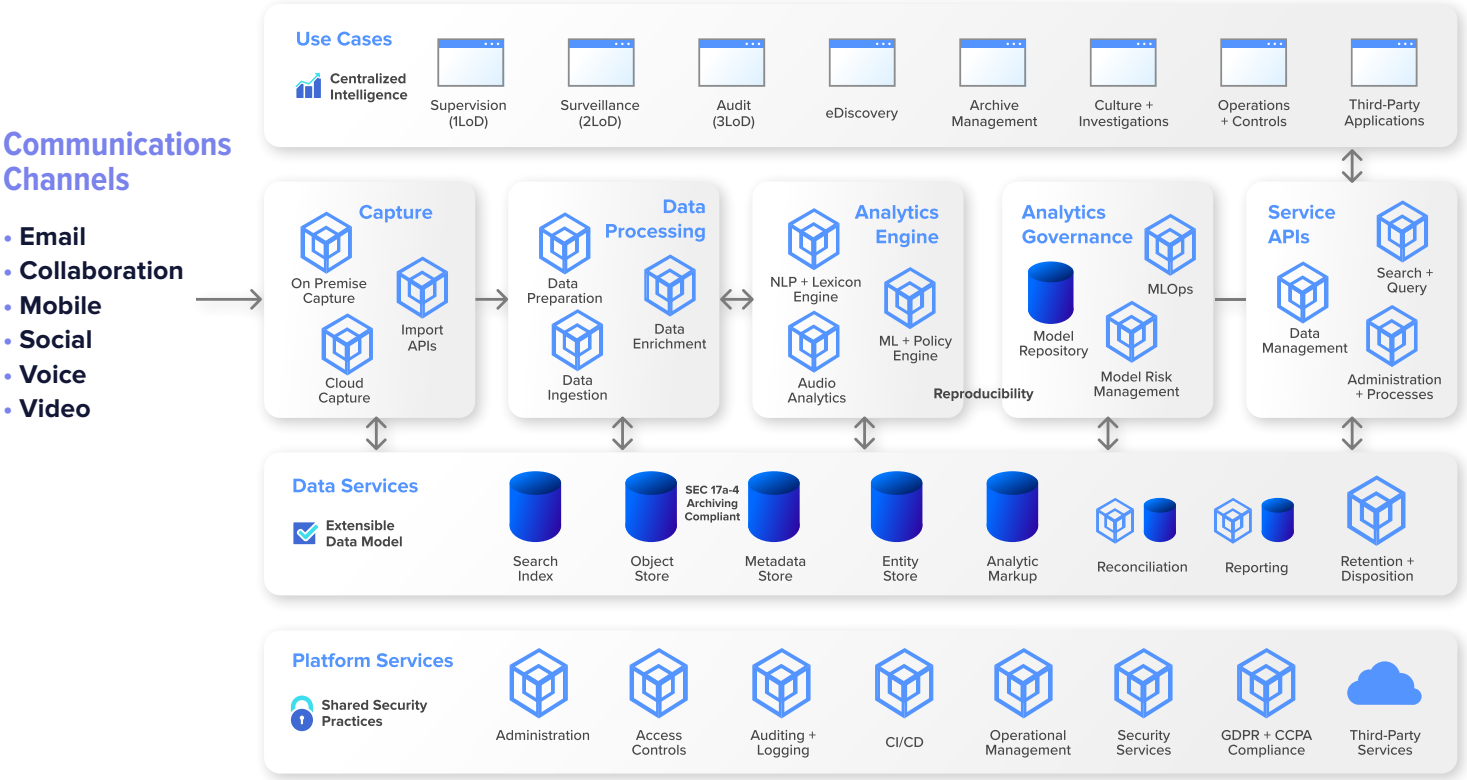
1. Communications data is captured from various communications channels.
2. The data is sent to the Communications Data Warehouse, where it is normalized, indexed and aggregated. Appropriate access controls and mechanisms are put into place. Regulated data is stored in a WORM-compliant manner.
3. Rule-based and machine learning analytics are applied to the data, and the findings are stored in the Communications Data Warehouse for downstream applications.
4. Applications enable key functions: alert reviews, investigations, reporting, actions and escalations, and internal and external audits.
5. Human judgments are passed back into the analytic system to improve the quality of the machine learning algorithms and automate additional work.



Reference Architecture

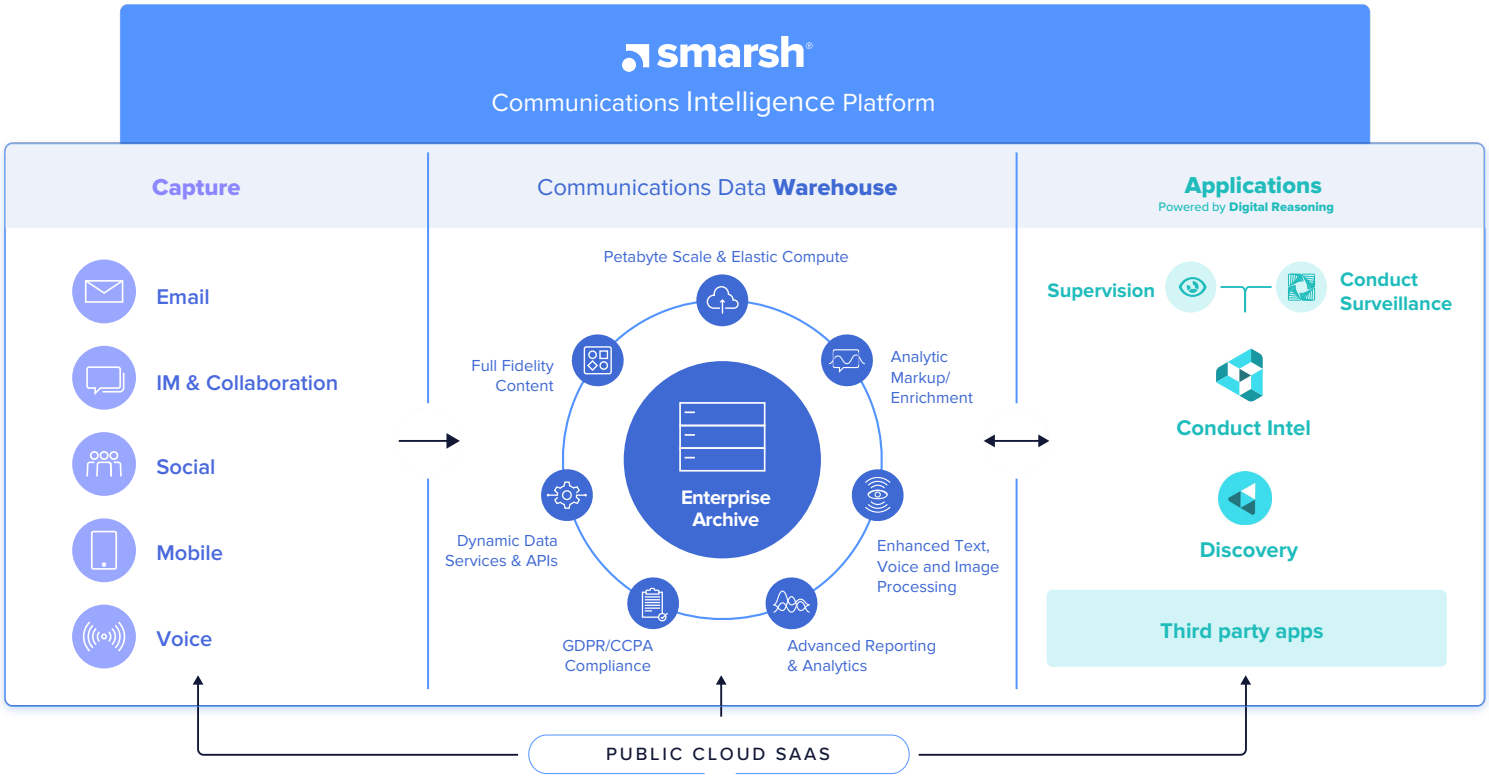
Top-level components can be further expanded to include common components within a Communications Intelligence technology stack. While the Smarsh offering includes many of the items listed below, they do not map directly to our offering. Rather, these are the services that we believe will emerge as common as the industry evolves.

Below we outline a reference architecture for a Communications Intelligence system. A description of the various services is available in Appendix A.



The Smarsh Communications Intelligence Platform:

The Communications Intelligence Platform is the first-of-its-kind SaaS platform that is AI-powered, cloud-native, and built to scale to meet communications data needs of modern global enterprises.



Capture
Smarsh captures all of the most popular email, mobile, social, IM & collaboration, video and voice tools used today. Retain and index important contextual details to speed up and improve supervision and e-discovery reviews.

Enterprise Archive
A cloud-native, contextually aware, extensible archive for global enterprises with complex security, data privacy and regulatory requirements.

Conduct Intel
The combination of two powerful, industry leading products, Smarsh Enterprise Supervision and Digital Reasoning Conduct Surveillance, enables companies to act on signals of misconduct in electronic communications.

Discovery
Collect, preserve, review and export electronic communications data on-demand to reduce the time and cost of e-discovery.

Guidelines for Ethical Communications Intelligence

While most of this paper focuses on the technologies that enable Communications Intelligence, it is important to consider and address the ethics of Communications Intelligence. Technology outlines what's possible while ethics guide what's appropriate.

More specifically, today's breakthroughs in AI/ML enable firms to create a safer and more productive workplace than ever before, because they have more information. But the benefits must be weighed against considerations for privacy, bias and employee well-being.

We have developed a set of guidelines for ethical Communications Intelligence, listed below. Appropriate usage will vary firm-to-firm, and we find that our customers benefit from coordination with their internal ethics officers. As always, we welcome feedback from our user community to improve our practices and the industry at large.

1. Communications Intelligence should thoughtfully create a safer, healthier workplace while respecting individual privacy.
2. Communications Intelligence should decrease rather than increase institutional biases. All reasonable efforts should be taken to remove bias and discrimination.
3. Communications Intelligence decisions should be auditable, consistent and transparent.

To achieve these guidelines, firms will need to maintain a commitment to Ethical AI, which focuses on Robustness, Explainability and Governance.

The importance of privacy

At Smarsh we strive to ensure that we are creating technologies that improve the greater good of our customers and their employees. In support of this goal, our products are built with a "privacy first" mindset:

- Data can be warehoused in appropriate geographies to comply with GDPR requirements
- Data can be dispositioned to delete communications involving a specific participant, enabling the "right to be forgotten"
- Our model training data works to remove personal information such as names, phone numbers, drivers' license numbers and much more
- Our model governance processes enable manual review of your firms' training data to correct for biases and to create fair process
- In the future, we plan to tag communications when they contain Personally Identifiable Information (PII) or Personal Health Information (PHI)

The Future of Communications Intelligence

It is predicted that by 2025, the total volume of data created and shared worldwide will reach 181 zettabytes — a 1068%[8] increase from 2015. That ten-year span has also seen the advent of widely popular communications tools such as Microsoft Teams and social media apps like Tik Tok, the emergence of cryptocurrencies as a viable financial product, retail investing applications, and fully remote or hybrid global enterprises.

The tools we use have changed, the financial markets have changed, the very structure of the corporate office has changed. As we look to the future, it's important to lay the foundation for your infrastructure now to stay prepared for shifting sands. Communications Intelligence as the basis for our platform is designed to be future proof. And we will continue to partner with the industry to shape this future.

8) https://www.statista.com/statistics/871513/worldwide-data-created/

Appendix A: Component descriptions

Capture

- **Cloud Capture** to collect communications across most written, spoken and visual communication channels (e.g., Email, Slack, Teams, Zoom).
- **On-Premise Capture**, when needed, to capture non-cloud communications (which often come from proprietary internal systems).
- **Import APIs** enable extensibility to cover non-vendor provided channels.

Data Processing and Analytics

1. **Data Preparation, Ingestion and Enrichment** includes traditional Extract, Transform and Load functions as well as the integration of person-centric metadata from other systems.
2. **ML, Policy, NLP and Lexicon Engines** combines rules, filters, natural language processing and other machine learning algorithms to extract key signals from data.
3. **Audio Analytics** includes transcription, speaker diarization, speaker identification and potentially other services such as language detection and paralinguistics.
4. **Model Repository** stores, versions and manages change control of machine learning models.
5. **ML Scenarios** used to derive insights and store them as analytic markup. The ML Scenarios leverage machine learning and rules-based logic to surface business critical signals.
6. **Scenario Management** tools enable the creation and governance of ML Scenarios.
7. **Policy Management** defines which Scenarios apply to which populations, and the corresponding workflows.
8. **MLOps and Model Risk Management Tools** monitor the operations and performance of machine learning scenarios.

Data Layer

- The **SEC 17a-4 Archive** provides WORM-compliant books and records storage.
- A **Communications Graph** connects conversations that span across communications.
- An **Entity Store** resolves individuals across communications handles and links them to appropriate departments and corporate structures.
- **Analytics Markup** stores "intelligence" such as analytic insights for reuse across eligible applications.
- **Reconciliation Services** ensure that the “data in equals the data out” to measure the “completeness” of your data processing.
- A **Reporting and Analytics** service packages intelligence and insights into value-added visualizations and time series (which can be sent downstream to trend and anomaly detection systems).
- **Retention and Disposition Data Management** tooling enables firms to dispose of relevant communications in accordance with regulations, and inspection of potential PII and PHI.

Platform, Services + APIs

- **Search and Query** services enable full-text search, attribute searches and advanced searches such as those with wildcards and Boolean operators.
- **APIs and Access Controls** enable data and insight consumers to connect into the data that they are entitled to see.
- A **Data Export** service enables fine-grained and access-controlled exports to other consumers of data.
- **Administration** tools govern user and access management similar to other technology offerings.

smarsh®

Smarsh enables companies to transform oversight into foresight by surfacing business-critical signals in more than 80 electronic communications channels. Regulated organizations of all sizes rely upon the Smarsh portfolio of cloud-native electronic communications capture, retention and oversight solutions to help them identify regulatory and reputational risks within their communications data before those risks become fines or headlines.

Smarsh serves a global client base spanning the top banks in North America, Europe and Asia, along with leading brokerage firms, insurers, and registered investment advisors and U.S. state and local government agencies. To discover more about the future of communications capture, archiving and oversight, visit www.smarsh.com.

US: 1-866-762-7742 | UK: +44 (0) 20 3608 1209 www.smarsh.com @SmarshInc SmarshInc Company/Smarsh