A 1LoD publication

in-Focus® Report

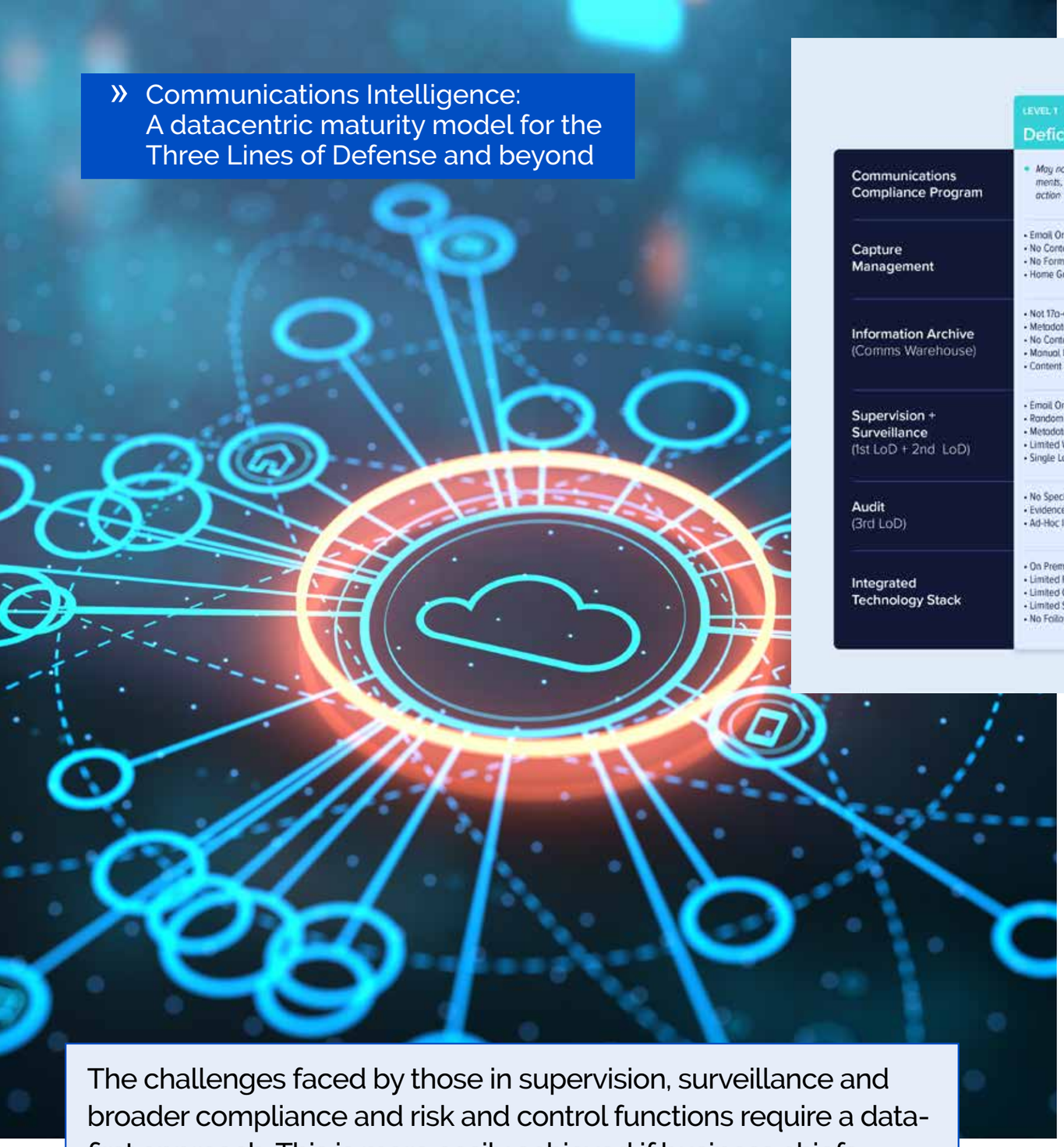
Commissioned by: smarsh®

Communications Intelligence:

A datacentric maturity model for the Three Lines of Defense and beyond

January 2022

» Communications Intelligence: A datacentric maturity model for the Three Lines of Defense and beyond



	LEVEL 1 Deficient	LEVEL 2 Developing	LEVEL 3 Operational	LEVEL 4 Advanced	LEVEL 5 Transformational
Communications Compliance Program	• *May not meet regulatory requirements, high risk of regulatory action*	• *Begins to achieve regulatory requirements, some risks surfaced via compliance program.*	• *Regulatory compliance. Program able to surface some true risks with high degree of noise.*	• *Program able to surface true risks with lower noise in predominant languages in eComms.*	• *Proactive risk detection across languages, cultures and mediums.*
Capture Management	• Email Only • No Content Richness • No Format Consistency • Home Grown	• Email + Collaboration • Email Threading • Inconsistent Formats • Multiple Vendor	• Most eComms + Mobile • Flattened EML Content • Flattened Standard Schema • Multiple Vendors	• eComms , Social, Basic Voice • Some Rich Content • Mixed Flat/Native Schema • Vendor Consolidation	• eComms, Social, Adv. Voice • Full-Fidelity Content • Standardized Schemas • Centralized Single Vendor
Information Archive (Comms Warehouse)	• Not 17a-4 Compliant • Metadata Search • No Content Richness • Manual Exports + Disposition • Content Only Storage	• 17a-4 Certified • Full Text Search • Email Threading • Exports, Time-Based Retention • Content Only Storage	• 17a-4 Certified • Advanced/Saved Search • Flattened EML Content • Exports, Role-Based Retention • Content and Manual Markup	• 17a-4 Certified • Entity/Faceted Search • Some Rich Content • Read APIs, Geo Retention • Content and Analytic Markup	• 17a-4 Certified • Multilingual Advanced Search • Full Fidelity Content • Write APIs, Privacy Retention • Automated Content Markup
Supervision + Surveillance (1st LoD + 2nd LoD)	• Email Only • Random Sampling Only • Metadata Search • Limited Workflow • Single Language Coverage	• Email + Collaboration • Basic Lexicon Policies • Full Text Search • L1/L2 Workflow • Multiple Languages Covered	• Most eComms + Mobile • Rich Lexicon Policies • Advanced/Saved Search • Review Queues/Escalations • Whitespace or Continuous Languages	• Most eComms + Basic Voice • Basic ML, Rich Lexicons • Entity/Faceted Search • Segregation of Review • Whitespace + Continuous Languages	• All eComms + Advanced Voice • Advanced Machine Learning • Multilingual Advanced Search • Case Mgmt + Work Allocation • Multilingual Technologies
Audit (3rd LoD)	• No Specialized Roles • Evidence of Supervision Report • Ad-Hoc Investigation Process	• No Specialized Roles • System Setup Audit Report • Search-Based Investigations	• General Admin Roles • Message Review Audit Report • Timeline-Based Investigations	• Internal Audit Role • Policy History Audit Report • Network-Based Investigations	• Internal/External Audit Roles • Automated Audit Reporting • Behavioral Anomalies
Integrated Technology Stack	• On Premise • Limited Recon/Identity Mgmt • Limited Observability • Limited Security Certifications • No Failover	• On Premise or Private Cloud • Some Recon/Identity Mgmt • Some Observability • Some Security Certifications • Periodic Backups	• Mixed Public / Private Cloud • Per App Recon/Identity Mgmt • Some Observability • Infrastructure + App Certified • Data Center Replication	• Single-Geo Public Cloud • Per App Recon/Identity Mgmt • Strong Observability • Per App Shared Security Model • Cloud Replication	• Multi-Geo Public Cloud • Unified Recon/Identity Mmt • Strong Observability • Cloud Redundancy

The challenges faced by those in supervision, surveillance and broader compliance and risk and control functions require a data-first approach. This is more easily achieved if business chiefs realize that they too benefit from the same tooling and data model. A new maturity matrix defining the journey to this target state is a significant first step.

Financial institutions worldwide are at a crossroads. Over the past 10 or more years, under increasing regulatory pressure, they have spent billions of dollars building and improving their management of non-financial risk. Adopting the Three Lines of Defense methodology, they have built extensive in-house teams, hired additional third-party service providers and both built and bought large and complex technology stacks to combat market abuse and misconduct, to prevent financial crime and to improve their cultures.

But the brute force approach to compliance and risk management they have employed is reaching its limits. For some, it has not even enabled them to achieve compliance with regulatory minima. For the compliant, it has simply become clear that their three lines of defense models cannot deliver the efficiency and effectiveness they need at a sustainable cost. Furthermore, the rules-based systems they are using do not deliver true risk management or mitigation – and in the case of financial crime prevention, they do not deliver enough useful intelligence to law enforcement nor do they significantly disrupt the criminals.

The more sophisticated institutions on both the buy- and the sell-side have turned to what is loosely termed 'artificial intelligence' (or AI) to improve their performance. These technologies in theory make possible the ingestion and comprehension of vast amounts of text-based information derived from e-communications channels or transcribed voice communications; and they hold out the promise of being able to analyze huge pools of data and metadata to identify previously unseen patterns of communication and behavior that may indicate misconduct. Some have already bought solutions; others have developed in-house tools in a 'skunkworks' environment.

But banks are finding that neither brute force nor smart technologies are delivering the improvements in efficiency and effectiveness that they need. Why? Because obsolete hardware and software, poorly designed data capture and aggregation models, and disparate internal data silos preclude the successful deployment of new analytics technology.

Across surveillance, financial crime, audit, ESG, cyber and elsewhere, leaders say the same thing: their core problem is data – data availability, data quality and data visibility.

From self-interest to best practice

The data problem is ultimately not simply a technology problem, it's an organizational problem. As Brian Cramer, CEO of Smarsh explains, "To address the problems most institutions face across the three lines of defense you need two things. First, you need the data infrastructure side. All the required electronic communications must be captured in such a way that it can either be transformed into usable data for machine learning models, or in a form that is already appropriate for a machine learning model to consume. Second, there is the analytics engine, the AI component itself, which derives the insights from the data. At most financial organizations today, these two efforts are independent of each other – siloed."

The question banks face is how to align the interests of the relevant stakeholders so that they can move away from the current situation, in which no single function has the leverage to move to a datacentric approach. This can only happen if the traditional silos – supervision, surveillance, audit, legal, HR, broader compliance, and the business itself – can be persuaded that while they all require different outputs and outcomes from technology, the tooling required to produce those outputs is the same.

Most critically, banks must find a way to operationalize an insight with which they are already familiar: a successful communications surveillance and compliance operation is, by definition, the custodian of a unique and vast trove of bank-wide data with uses far beyond compliance.

This dataset, aggregated and analyzed appropriately, is a source of insights that not only provide solutions for today's rules-based regulations, but also for tomorrow's risk-based approaches, as well as future regulatory extensions around ESG, digital assets, cyber and resilience.

Most significantly, such a dataset also generates business intelligence with the potential to directly benefit the P&L of divisions who will have up till now seen the data needs of the three lines of defense as just a compliance-related cost to be driven down as aggressively as possible.

As one recently retired head of surveillance at a global bank says, "Once you have all the comms data in one place, then you can look not just for market abuse but also at financial crime, ESG and culture. Boards and investors increasingly want metrics on what's happening in these areas so I would encourage Smarsh to push on with this. The model is clearly correct in its implication that getting your data in order is 90% of what matters. A single source of this information is valuable not just to the regulatory compliance side but also to the commercial side."

If those in charge of compliance and surveillance and those who run businesses accept that the fundamental tools they require are the same, then banks can escape the cycle in which individual functions end up creating their own band-aid solutions, buying or developing their own analytics solutions.

Bridging the gap between compliance and the business

To help conceptualize the links between these seemingly different needs, Smarsh proposes a maturity model that describes progress towards best practice as a journey, and gives milestones and KPIs for that journey, around both data infrastructure (broken into capture management, information archive, and the integrated technology stack) and the three lines of defense, (to represent the first, second line and audit, the third line).

As Cramer explains, "We saw a need for this because every customer we talked to had a different answer and struggled to articulate what their destination was. The model allows organizations to step back from everyday compliance to see where they are and where they should be going. And if they get this right, they will not only begin to solve their efficiency and effectiveness problems, but they will also build a new intelligence resource with the power to significantly enhance their businesses. And what's really interesting is that all of the technology exists today to do this, it's really about the leadership and the vision of the compliance and risk leaders at the banks to get there."

The matrix is not designed to mimic banks' current reporting lines. Clearly, for example, audit, second line surveillance and first line supervision do not typically work together on data capture. Instead, the matrix describes the datacentric architecture required to deliver the right outcomes for each function and the actions within that architecture that each function can take in the journey towards best practice.

In the face of the many data, technology and other challenges faced by the three lines, the question for most banks is 'what next?'. What should they do next to improve their compliance, to help them move from a rules-based to a risk-based approach or just to help them get rid of false positives? What do they need to do in terms of people, process and technology to help with first/second line convergence, to improve effectiveness across conduct, culture, market abuse and financial crime?

To make these decisions, which may involve several simultaneous initiatives, and to help build a business case for the resources required, banks need a roadmap of options, and potentially a benchmark with their peers.

The Smarsh maturity model is a way of breaking down the problems into separately defined streams of initiatives that can be practically addressed and which together define the infrastructure required for best practice communications data aggregation and analytics. Relative progress in each of these streams can be used to benchmark institutions against each other or against internal goals.

Without this kind of formal framework, it is very difficult for organizations to pursue coherent strategies around data. And in general, banks tend to be at very different stages of development from area to area. For example, one top 10 bank in the US captures 54 different types of communications for regulated employees and operates at 1.8 million transactions a day. In terms of data infrastructure it is extremely advanced. But this same institution uses no true machine learning or AI-driven analytics across that communications data.

On the other hand, it is common to find investment banks whose ability to provide analytical insights, whether they're risk- or compliance- or business-focused, is extremely advanced, but they run those analytics on a small subset of the data that they have and that would be relevant to their own specific use because they don't have the ability to capture and provide data in a scalable way to serve up that broader analytical functionality.

And there are institutions with the most sophisticated analytics, limited by the fact that they maintain their communications data on-premises. This constraint means that they capture only six to seven types of communications data while the employees that they're surveilling are using up to 50 different communications platforms.

These institutions' piecemeal approach to the problems of data and analytics leaves them with holes either in their data infrastructure or their analytics capabilities.

So how does the maturity model help them?

As Cramer explains: "We said, let's use a framework that already exists which led us to surveillance, compliance and so on and from there look at how to think about the role technology plays in each of those."

In other words, the model looks at the workflows that ultimately produce the outcomes banks want and is agnostic as to where individual banks carry out those functions (e.g. in the first or second lines of defense).

So, the data infrastructure layer is represented by streams that define capture of communications, archiving (storage, search and retrieval) and the technology stack upon which these functions are built. And the analytics layer is represented by maturity in the key surveillance functions, in audit and in the overall compliance program. Each of the streams is then broken down into five stages of development:

Compliance program:

The pandemic simply accelerated the adoption of multiple communications and collaboration channels, and the evolution of email into a legacy communications format. Regulators and compliance professionals have struggled with preventing these new channels becoming a haven for misconduct and the market needs a set of 'guard rails' to guide its development. The model acknowledges the deficient and developing stages, because many organizations do still struggle with basic compliance tasks. But it also defines a pathway to a fully holistic program in which proactive risk detection across all required languages, cultures and media is attainable.

Capture management:

A fundamental question for all organizations is, 'are you capturing the data in a way that allows you to derive value from it with machines?' This means capturing the data in its original context. Cramer explains, "So if it's a threaded teams discussion, can you reproduce it in that threaded manner and understand who said what and when, who joined the conversation and when, who left the conversation and when, and so on."

The Smarsh model envisages a vendor-neutral taxonomy and its own platform is an example of what modern, sophisticated data capture technology can do, providing multiple APIs and simple mechanisms for ingesting all forms of communications data and preserving them in their native format and context.

Information archive:

Once data is captured it needs to be stored in a secure, searchable and retrievable communications warehouse. One of the key issues here is immutability (in the US, SEC Rule 17a-4 compliance is key), so data must not be able to be changed in this archive without a complete record of any changes. And this is where you would think in terms of the traditional repository archive foundation.

The model envisages a path from failure to comply with 17a-4 all the way through to a fully operational repository service with data and metadata search, multilingual, entity-based search, and automated content markup. As data types and formats continue to proliferate, this data warehousing aspect of the compliance effort will become ever more critical.

Surveillance and analytics:

With the data captured, maturity is then indicated by an organization's ability to extract insights from it. As Cramer says, "To generate real intelligence you need to be applying machine learning and natural language processing to the data. That is the only way to get through that volume without adding armies of people which, as we all know, is unsustainable for any kind of risk function. So, you're almost left with no choice but to start to adopt machine learning, to deal with the volumes that you're getting today, to find as much risk as you can, as efficiently as you can."

The model recognizes that this work falls into two core categories, both executed in the first and second lines of defense.

First, banks have to surveil a series of defined populations through the lens of a series of defined risks. A defined population might be regulated employees; defined risk means a known set of policies set up around those employees driven, mostly, by specific regulatory requirements. This is the 'bread and butter' of surveillance – sifting through alerts generated by rules-based systems and, in the case of e-communications, lexicons – and escalating the very few that rise out of a huge volume of false positives.

"To generate real intelligence you need to be applying machine learning and natural language processing to the data. That is the only way to get through that volume without adding armies of people which, as we all know, is unsustainable for any kind of risk function."

"So, this is a monotonous, serial workflow in which human teams are reviewing communications to the tune of thousands per week, hundreds of thousands per year. They started off just looking at email and looking for needles in haystacks – something that indicated spoofing or money laundering or whatever. And now that's compounded across more markets, more regulations and types of misconduct, and with many more comms channels – zoom, teams, slack – which make that workflow even more challenging. And AI and ML are now inserted in there to make it feasible to find that needle," explains Cramer.

Second, banks also have an additional need to more proactively search for unknown risks and problems that fall outside of specific, easily codified policies.

Here the model acknowledges that BAU – largely driven by regulations – starts with that high-volume process and that in most cases this is a mature technology and set of procedures. It can still be improved – for example, by moving from random sampling to more continuous monitoring of full data sets. That again requires advanced machine learning and multilingual search, and the application of those newer technologies to the basic high-volume process leads to the virtuous circle of reduced false positives and better identification of true positives.

Integrated technology stack:

The evolution of compliance, supervision, surveillance and audit, the creation of a new data infrastructure and warehousing, and the application of next-generation analytics to that data, all rely upon a core hardware and software layer and architecture. And that layer has to cope with massive scale and complexity.

As Cramer says, "If you look at the top 50 banks, say, the volume of communications data that they are ingesting on a daily basis and the amount of data that they're carrying around from the last 10 years is a petabyte problem. So, they need something that scales and is able to scale at that rate. And there's really only one answer today, and that is public cloud infrastructure. It is no longer really feasible to keep all that data on-premise. There are many issues around cloud, third-party vendors, proprietary datasets and so on, but for all organizations there gets to a point where there is a strategic decision to be made: what you do with your data, where do you keep it and how do you protect it?"

The Smarsh model assumes that cloud adoption is inevitable and beneficial, and that public cloud (not private or hybrid) is the most likely end-state. "As communications proliferate, as the volume of data accumulates and becomes more challenging to grapple with multiple formats, multiple stakeholders, different models, different lexicons, multiple sets of indices and multiple geographies, you encounter more and more complexity and you need more processing power to interrogate that data," says Cramer. "At that point, it becomes necessary to look at the public cloud approach and at least have elements of it in your information architecture."

The marketplace is already proving these assertions correct. The world's banks, from Tier 1 global to country champions, are adopting public cloud. And, as one former senior regulator points out, "I am a hundred percent behind [Smarsh] on storing this data in the public cloud. FINRA has had significant amounts of its data in AWS (Amazon Web Services) since at least 2014 and the FCA has been using the public cloud since 2015. Despite this, some banks have been really reluctant to do this, and I think they need to reflect on their reluctance in the light of the fact that two globally significant regulators have been doing it for years. There's no way that they actually could do what they do with the data without using the cloud."

Regardless of final technology destination, organizations will need solutions that address issues such as data copies, outages, redundancy, observability, the ability to react to infrastructure conditions, reporting, security, authentication, and identity management. The integrated technology stack is the foundation of any sophisticated communications surveillance and compliance operation.

From compliance to the business

The model, and the process it describes, starts from the idea of doing compliance better by optimizing data, technology and process in each of the core compliance building blocks. However, communications data contains very significant value in terms of deriving insights into clients, markets, strategies and business opportunities, providing firms with a revenue incentive as well as a compliance incentive to look at where they fall along the spectrum of maturity in terms of leveraging their communications data as a complete asset.

Smarsh calls this 'Communications Intelligence' because, like Business Intelligence before it, it brings together different sets of data from siloed parts of the business to produce insights that could not otherwise be produced. "Aggregation of data overlaid with sophisticated analytics is the key to extracting intelligence. Communications Intelligence starts with compliance because they're already the largest consumer of communications data – far beyond any business user. But there are many more opportunities that go beyond compliance if you can get data capture, the technology stack and the machine learning analytics right," says Cramer.

Compliance is the driver, but the ultimate opportunity is much greater and sits across all three lines of defense. The model allows organizations to understand where the challenges lie in creating a more holistic approach to using communications data at scale to find risk. The smartest organizations understand that they can use this model to accomplish more sophisticated conduct objectives that have nothing to do with regulations or regulated populations. They can identify cultural issues before they become a problem. They can identify successful or unsuccessful traits in businesses or individuals. And they can proactively surface risks and opportunities with their clients.

The model starts with a maturity model for the people, processes and technology around core compliance; it ends with Communications Intelligence.

As one former surveillance head says, "I see very significant value in a standardized maturity model for e-comms. It will greatly help financial sector conversations about the future of e-comms, which currently are disjointed, with financial institutions comparing apples with oranges and each vendor selling their USP in language particular to them."

If you look at the top 50 banks, say, the volume of communications data that they are ingesting on a daily basis and the amount of data that they're carrying around from the last 10 years is a petabyte problem. So, they need something that scales and is able to scale at that rate.

Hear more from Brian Cramer about the maturity model in this podcast

"I see a lot in here I agree with. [The model] is a good way of looking at things and I would like to be able to take it to the committees that we have where all the stakeholders around this data make these kinds of decisions," says one head of global market surveillance strategy.