GDPR compliance steps, and impacts on MDM thinking

The impacts of the General Data Privacy Regulations are tremendous and cannot be overstated for any business that deals in personal or private data. In large enterprises, the subject of Master Data Management is often a 3rd rail, or considered a sure path to career death due to the complexity of the topic. Done properly the benefits are tremendous, but the ownership of information at the functional level is so often a barrier to success that it has become generally understood that these efforts are rarely fully successful.

Now we enter the world of GDPR. The regulation in principle is simple. It can be generalised to say: If you collect any personal information (broadly defined) you must make the information owner aware of what you collect, what you do with it and how to get that information back or purge it.

While it sounds simple in application, the complexity of the average enterprise information ecosystem makes this horrendously challenging. In the pharmaceutical industry where I currently work, patient information is collected in the course of a number of activities. As a part of that, various systems and reports are impacted and the process gets exponentially more complicated when biospecimens (samples of tissue, blood, etc…) are concerned. All samples must have consent for use documented, and a traceability through the full lifecycle. When data is “blinded” or anonymized, this adds even more complexity as the ramifications of “unblinding” data are significant to the integrity of the overall trial process. Beyond patient data, there is data on clinicians, research organizations, collaborators and more.

My recommendations I have made, and I suggest to clients include the following:

1. A comprehensive audit of all Personally Identifiable Information collected or generated in the course of business. This should result in a list of data attributes, source of collection, systems and processes impacted
2. For all data collected, where mastering or simplification of process is feasible, a plan must be defined and should be mapped into an overall value stream analysis for inclusion in a portfolio of work focused on GDPR compliance.
3. All data once mapped, must be included in a reference that allows comprehensive aggregation of the record – unifying the person to a single view. This requirement is where the value of mastering comes to the top. Without the ability to view a person in a record, the next step is exponentially more challenging.
4. The Data Privacy Officer for the company and if applicable, related functions, must be pulled in to evaluate the plans and approach, supplementing the data record or approach as needed.
5. The next step in compliance once the data is clearly mapped and understood is an evaluation of the impact of “providing and purging” as each data owner has the right to request at any time a comprehensive export of all data collected about them. Correspondingly, they can also request that all personal data be purged. This becomes particularly challenging when data has cross dependencies and is a part of a larger analysis.
6. The evaluation of the provide and purge process must result in a documented program and process to answer the question of what is the impact of a request. How will we mitigate the risk and what needs to change about our processes to accommodate these needs. This starts with the evaluation, moves to a

documented plan, and then must be accompanied by at a minimum, a tabletop exercise of the multiple scenarios identified.

This is a costly effort, from a time and resources perspective. As an enterprise, leaders must be aware of the risk of non compliance relative to the cost of compliance.

GDPR compliance steps, and impacts on MDM thinking

The impacts of the General Data Privacy Regulations are tremendous and cannot be overstated for any business that deals in personal or private data. In large enterprises, the subject of Master Data Management is often a 3rd rail, or considered a sure path to career death due to the complexity of the topic. Done properly the benefits are tremendous, but the ownership of information at the functional level is so often a barrier to success that it has become generally understood that these efforts are rarely fully successful.

Now we enter the world of GDPR. The regulation in principle is simple. It can be generalised to say: If you collect any personal information (broadly defined) you must make the information owner aware of what you collect, what you do with it and how to get that information back or purge it.

While it sounds simple in application, the complexity of the average enterprise information ecosystem makes this horrendously challenging. In the pharmaceutical industry where I currently work, patient information is collected in the course of a number of activities. As a part of that, various systems and reports are impacted and the process gets exponentially more complicated when biospecimens (samples of tissue, blood, etc…) are concerned. All samples must have consent for use documented, and a traceability through the full lifecycle. When data is “blinded” or anonymized, this adds even more complexity as the ramifications of “unblinding” data are significant to the integrity of the overall trial process. Beyond patient data, there is data on clinicians, research organizations, collaborators and more.

My recommendations I have made, and I suggest to clients include the following:

1. A comprehensive audit of all Personally Identifiable Information collected or generated in the course of business. This should result in a list of data attributes, source of collection, systems and processes impacted
2. For all data collected, where mastering or simplification of process is feasible, a plan must be defined and should be mapped into an overall value stream analysis for inclusion in a portfolio of work focused on GDPR compliance.
3. All data once mapped, must be included in a reference that allows comprehensive aggregation of the record – unifying the person to a single view. This requirement is where the value of mastering comes to the top. Without the ability to view a person in a record, the next step is exponentially more challenging.

4. The Data Privacy Officer for the company and if applicable, related functions, must be pulled in to evaluate the plans and approach, supplementing the data record or approach as needed.
5. The next step in compliance once the data is clearly mapped and understood is an evaluation of the impact of "providing and purging" as each data owner has the right to request at any time a comprehensive export of all data collected about them. Correspondingly, they can also request that all personal data be purged. This becomes particularly challenging when data has cross dependencies and is a part of a larger analysis.
6. The evaluation of the provide and purge process must result in a documented program and process to answer the question of what is the impact of a request. How will we mitigate the risk and what needs to change about our processes to accommodate these needs. This starts with the evaluation, moves to a documented plan, and then must be accompanied by at a minimum, a tabletop exercise of the multiple scenarios identified.

This is a costly effort, from a time and resources perspective. As an enterprise, leaders must be aware of the risk of non compliance relative to the cost of compliance.

Data Privacy – coming to a state near you….?

Actions in California this past year led by a real estate developer named Alastair Mactaggart have the potential to change the privacy game in the United States as a whole. I have included the legislation as an embedded pdf below, and you can link to the California government site here to read the legislation as well.

NPR wrote an article on this topic, and quoted Mactaggart.

> "These giant corporations know absolutely everything about you, and you have no rights," he said in an interview earlier this week outside the state Capitol. "I thought, oh, I'd like to find out about what these companies know about me. Then I thought, well, someone should do something about that."
> Eventually, Mactaggart decided, "maybe I'm someone."
>
> NPR – HEARD ON MORNING EDITION QUOTING ALASTAIR MACTAGGART

The fact that Mr. Mactaggart was able to get the legislative support to push past the tech giants and others attempting to block this measure is no small feat. Now the big thing to watch for is how this drives the national policy makers. The US business market is at a real risk of patchwork privacy regulations, and the California laws could be the first in a series of dominoes that represent recognition of consumers right to own their information. As states grapple with this reality, and build on the European GDPR model and now the California laws, the national government ignores this pending wave of regulation at both its own peril, and the peril of the US

business economy.

I encourage you to read the California legislation and think about the impact to your life, and the business you conduct. I work in a large international corporation, and as a part of that, GDPR has become very real for myself and my teams. While privacy should always be a primary concern, this is an opportunity to join the discussion and embrace the fact that regulation is coming, and rather than wait or ignore it, when possible make our voices heard in support of national level legislation. Localized and state level variations in regulations that impact business across state lines is costly and complicated to comply with and to enforce. A comprehensive national approach modeling on the learnings of our EU colleagues and now the California legislation, is our best bet for stability through this transition process.

Is there truth on the internet?

The growing sophistication of our technology platforms, and the related reliance on these tools for our social fabric, is creating an environment where the truth is increasingly difficult to discern. At first blush, it would seem this social web makes truth harder to conceal, as you cannot hide from the cameras that are everywhere, and the instant posting of twitter or other platforms. While this "instant sharing" is true, it is also true for "fake news" or intentional disinformation. I read a recent white paper by Senate Intelligence Vice Chairman Mark Warner (D-VA) and the line of thinking is both alarming and disturbing. Reading the paper, I am taken to a rather famous quote from George Orwell's 1984:

> "Who controls the past controls the future. Who controls the present controls the past."
>
> — GEORGE ORWELL, 1984

We are rapidly approaching this inflection point where government and those with deep enough pockets, combined with access to influence for social media platforms or channels, will be able to control the present through the news channels. This facilitates rewriting events to their narrative. Once this happens, control of opinion begins and the future is shifted. Not to go too far down the Orwellian path, but this whole line of thinking is not wholly without merit. Read the paper from Mr. Warner, and see what your conclusions are. I am firmly against government overreach and control of private data. That being said, there does need to be a way to create a regulatory oversight for the proliferation of false narratives. The question then becomes, who polices the authority, because the ownership of the truth is absolute power in the end.

Multi-Cloud Service Delivery

As I have been exploring the maturing environment of cloud services, I am regularly struck by the richness of the environments and the dramatic shift to “getting it done” with microservices, versus the legacy thinking of stack based development. There is much to dig into from an interoperability, scaling, global security model and more, but at present, the top three players in the space are offering a broad array of options that are sparking my thinking across a range of options and need spaces.

1. AWS (Amazon Web Services)
2. AZURE(Microsoft Cloud Services)
3. Google Cloud Functions

The next level of maturity is an established pattern for integration, that uses global security models to facilitate interop, with a common set of controls that sit on top and are referenced across all platforms and data stacks. Getting to the granular, element level in the data lake, secured by role and user is critical in the emerging privacy world. There is a clear need to have the capability to have a single world view of a person, or a resource across these platforms, abstracting the security model in a scalable way for both development and user engagement.

I am seeing articles pointing to this general thinking, but still not satisfied with a common “glue” or abstraction layer for these unified visions. I look forward to seeing this emerge, and being a part of that solution to the extent I am able.

GDPR – General Data Protection Regulation

The EU Regulations around data privacy and protection are emerging, and as they do, the initial rulings are in

effect. The below excerpt from the EU site references what is considered personal data – and specifically that which has been anonymised, or otherwise obfuscated. Note that even if the data as it sits is non-identifiable, if it can be combined to become identifiable, it falls under the guidelines of the regulation.

> The principles of data protection should apply to any information concerning an identified or identifiable natural person. Personal data which have undergone pseudonymisation, **which could be attributed to a natural person by the use of additional information should be considered to be information on an identifiable natural person**. To determine whether a natural person is identifiable, account should be taken of all the means reasonably likely to be used, such as singling out, either by the controller or by another person to identify the natural person directly or indirectly. To ascertain whether means are reasonably likely to be used to identify the natural person, account should be taken of all objective factors, such as the costs of and the amount of time required for identification, taking into consideration the available technology at the time of the processing and technological developments.
>
> The principles of data protection should therefore not apply to anonymous information, namely information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable. This Regulation does not therefore concern the processing of such anonymous information, including for statistical or research purposes.

This bit of the guideline is of particular interest to me, as part of what I see on a regular basis is the attempts to understand for each data related initiative my team undertakes, how to ascertain the potential impact of this regulation, and if it is applicable. This bit of text certainly makes it broadly applicable, and it seems good data hygiene is generally to make the assumption that any global system should plan to follow the general guidelines laid out by the regulations.

The somewhat complicating factor is that the regulations as of this writing are not in final form, and the penalties for non compliance are not trivial.

In my searching for more information on this topic, I came across a decent summary of the work – linked here. This site is not an official EU government site, but rather a vendor partnership education site. That being said, I think they do an admirable job of simplifying the regulation to language that the layperson can digest and use

to better prepare for compliance.

Ref Links:

- Vendor site with summary: https://eugdpr.org/the-regulation/
- Text of regulation, as of this writing (In English): https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679&from=EN
- EU site with Regulation and multi-language support: https://eur-lex.europa.eu/eli/reg/2016/679/oj