dataiku

FIVE ESSENTIAL PILLARS OF BIG DATA GDPR COMPLIANCE

The Path to Compliance Through Data Governance

WHITE PAPER

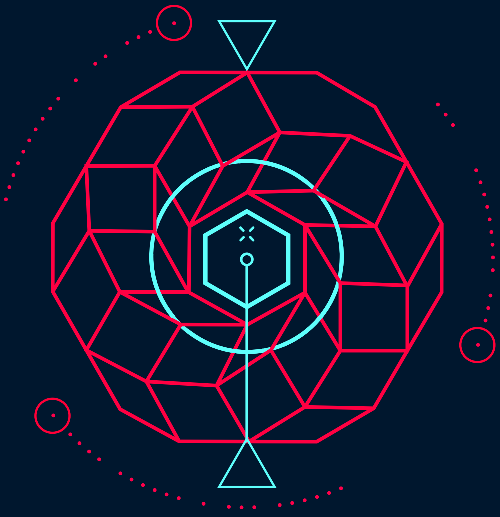


TABLE OF CONTENTS

INTRODUCTION AND BACKGROUND

About

This white paper includes a short introduction to the EU General Data Protection Regulation (GDPR) as well as tips and suggestions on how data governance and communication/ collaboration can aid compliance. It is not intended to be legal advice or all-encompassing, but rather a non-comprehensive look at the changes it will bring for teams across an organization.

Key Terms

This white paper will use the following terms, which are defined in both the current Data Protection Directive and also in GDPR. We've included simplified definitions here for easy reference:

PERSONAL DATA

Any information related to a human being (or data subject) that can be used to directly or indirectly identify that person. For example: name, photos, email addresses, bank details, posts on social networking websites, medical information, IP addresses, etc.

DATA SUBJECT

A human being on whom personal data is being collected.

SENSITIVE DATA

A special category of personal data (including personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, trade-union membership, and data concerning health or sex life) to which additional protections apply.

DATA CONTROLLER

An entity that determines the purposes, conditions, and means of the processing of personal data.

DATA PROCESSOR

An entity that processes personal data on behalf of the controller (e.g., cloud and datacenter providers).

Background

The Data Protection Directive (formally known as Directive 95/46/EC) has governed personal data protection in the European Union since its adoption in 1995. It's important to note that the Data Protection Directive is ***not a regulation***; it is a set of goals or guidelines relating to personal data (especially processing, using, or exchanging it) that individual member countries implemented separately. A regulation, by contrast, is a binding legislative act. As a result of this nuance, data protection legislation across the EU has been fragmented, arguably weakening its authority and resulting in legal uncertainty.

Due to the legal gray area and uncertainty of the previous Data Protection Directive, as well as the quickly changing data landscape, **the European Parliament developed the GDPR throughout 2015 and 2016 and adopted the GDPR on April 14, 2016.**

May 25, 2018

Enforcement

The GDPR will be officially enforced beginning on **May 25, 2018.** Ideally, systems and processes to support GDPR changes should be in place well before this deadline.

GDPR KEY TAKEAWAYS

GDPR: The Cliff Notes

The GDPR is thorough and should be reviewed closely by all affected organizations to ensure compliance. But at a high-level, the most important points to be aware of surrounding GDPR are as follows:

Application

The GDPR applies to any company (regardless of their location, size, and sector) processing the personal data of people residing in the EU. For example, a US-based company processing the personal data within the United States of EU citizens is required to comply.

Responsibility

Under GDPR, both data controllers and processors must comply with the legislation. Under the previous/current Data Protection Directive, only data controllers were held liable for data protection compliance, not data processors.

Penalties

With a maximum fine of up to 4 percent of annual global turnover or €20 million (whichever is greater), penalties for non-compliance are steep.

Consent

Under GDPR, companies will no longer be able to use long, illegible terms and conditions full of legalese; consent for collection and use of personal data must be in plain language and detail the purpose of data processing.

Data Breaches

Increased regulation surrounding the disclosure of data breaches; specifically, much quicker reporting is required (within 72 hours).

Data Subjects' Rights

EU data subjects will have expanded rights when it comes to data protection, including the right to be forgotten (have their data erased), the right to access (obtain information about exactly what data is being processed where and for what purpose), and the right to data portability (receive a copy of the personal data concerning them). Citizens now also have the right to question and fight decisions that affect them that have been made on a purely algorithmic basis.

Privacy by Design

It will be a legal requirement to consider data privacy on the onset of all projects and initiatives, not as an afterthought.

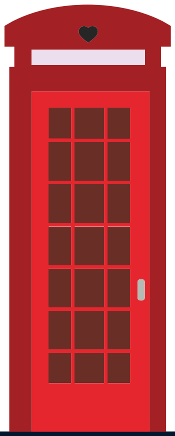
Data Protection Officer (DPO) Appointment

Controllers and processors whose core business is regular and systematic monitoring of data subjects on a large scale or who deal with special categories of data will be required to appoint a DPO. The DPO may be appointed from within, hired, or contracted, but (among other specific requirements) (s)he must be an expert on data protection law and practices.

GDPR OUTSIDE THE EU

Why Should I Care About GDPR?

United Kingdom



The GDPR will go into effect in May 2018, but by contrast, the United Kingdom is not likely to leave the EU until around May of 2019; so in fact, **there will almost certainly be some overlap between when the GDPR goes into effect and the United Kingdom's time in the EU**. Given the high fines, banking on non-compliance with Brexit impending is a huge risk, and regulations aside, putting good data governance practices in place is something businesses will benefit from in other ways anyway (including gaining customer trust).

Furthermore, companies based in the United Kingdom with have to comply regardless of Brexit if they have customers in the EU. And even for businesses strictly limited to the United Kingdom, it's important to note that the UK government has indicated it will implement equivalent or alternative legal mechanisms that will be largely similar to GDPR[1].

United States

The 2017 GDPR Preparedness Pulse Survey[2] conducted by PricewaterhouseCoopers (PwC) polled C-suite executives from large American multinationals and showed that **US companies are overwhelmingly aware of, and concerned with, GDPR regulations**. Over half of survey respondents cited GDPR as a "top" priority, and 38 percent named it "among" their top priorities. And rightfully so given that fines are applicable to US businesses as well and that the new regulations are relatively complicated and will require significant preparation, not just an afterthought.

It's easy to assume that just because your company doesn't have a physical or significant presence in the EU that the GDPR doesn't apply, but the text of the regulation makes it clear that any interaction with EU consumer data brings your company under its jurisdiction. If your website simply collects data on EU citizens, your company must comply. In short, GDPR applies to any enterprise in the world that targets the European market in offering goods or services or profiles European citizens.

Still Blurred

Lots of specifics surrounding exactly how the GDPR will work with Brexit as well as with cross-border data transfer regulations both in the United States and in other countries around the globe are still unclear. Officials and business leaders expect details and additional guidance to unfold in the coming months.

Given this uncertainty, the best course of action is not to ignore GDPR and wait for concrete next steps. Rather, it is to keep GDPR top of mind and begin preparations by implementing agile solutions that will begin to address the biggest and broadest provisions of the regulations. By starting to prepare but prioritizing flexibility, you can ensure your business doesn't get behind in its efforts while also not implementing any changes that back the company into a corner and prove to be wasted efforts in case of change.

Any business facing these uncertainties should also stay on top of the latest trends and work closely with regulatory agencies if possible to ensure compliance despite unsteady ground.

> *"No legislation rivals the potential global impact of the [GDPR]. The new law will usher in cascading privacy demands that will require a renewed focus on data privacy for US companies that offer goods and services to EU citizens," said Jay Cline, PwC's US Privacy Leader. "**American multinationals that have not taken significant steps to prepare for GDPR are already behind their peers.**"*

(1) GDPR FAQs." EU GDPR Organization, 26 April 2017, http://www.eugdpr.org/gdpr-faqs.html

(2) GDPR Preparedness Pulse Survey." PwC, 23 Jan. 2017, http://www.pwc.com/us/en/press-releases/2017/pwc-gdpr-compliance-press-release.html

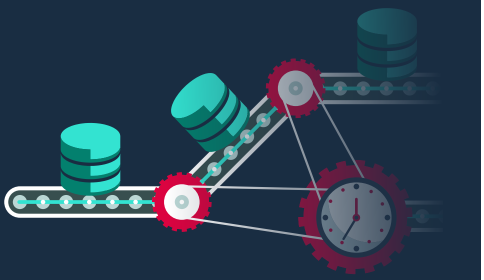
IMMEDIATE CHALLENGES

The GDPR brings about several major challenges that businesses will need to address before the enforcement deadline. The most daunting of these challenges surround data governance, organizational structure, and cross-team collaboration, including:

1 Determining where personal data is stored across multiple different (potentially siloed) data sources.

2 Aligning everyone across the company (including IT, marketing, customer support, and data teams) on new policies and execution of any changes.

3 Putting processes in place to accommodate requests from data subjects and ensuring all teams can execute on processes in a timely matter.



4 Ensuring proper data governance, security, and monitoring are in place in case of audit.

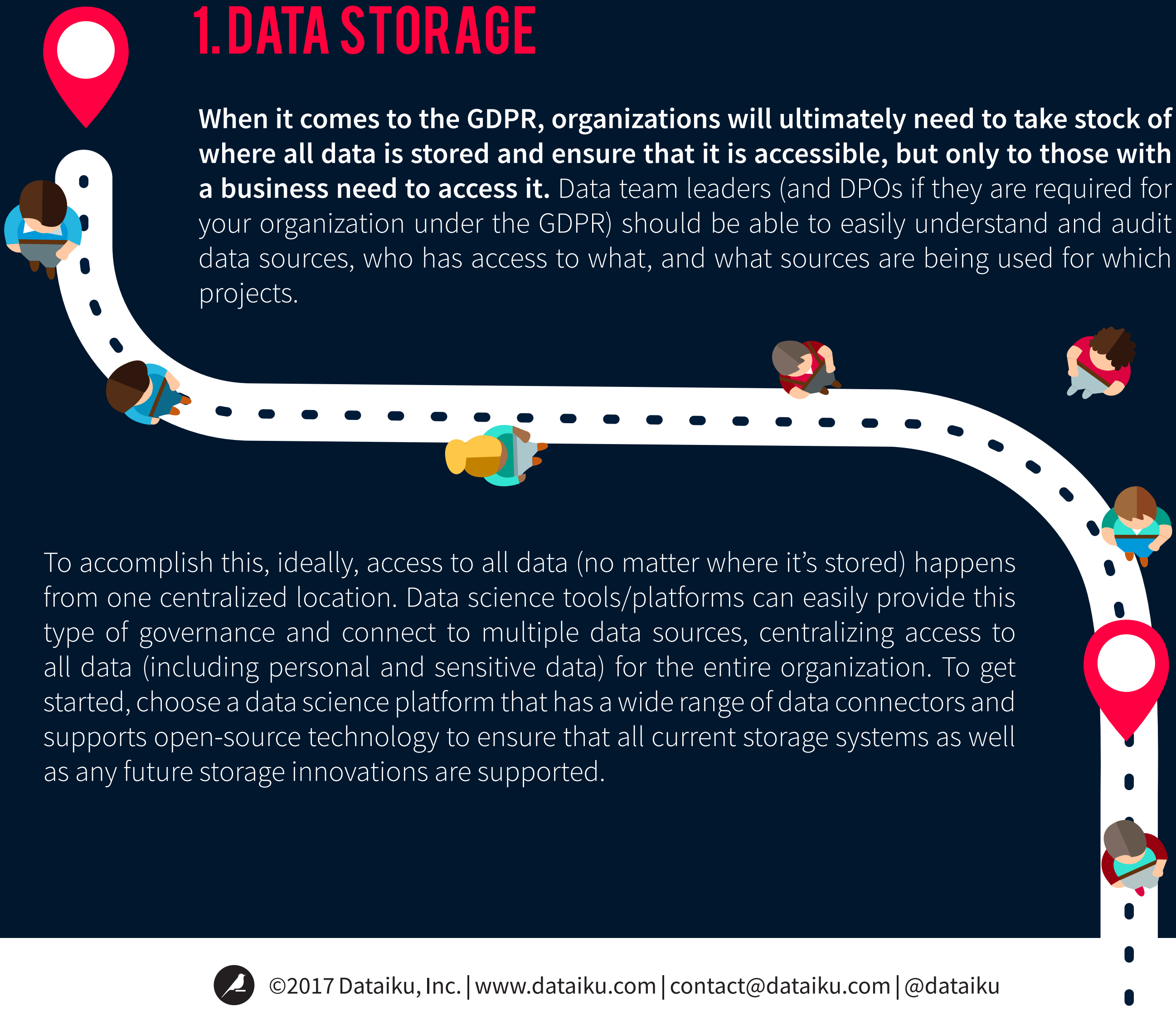
5 Implementing agile solutions that keep your operations flexible and easily adaptable to change.

Where to Begin?

The changes in GDPR will certainly require shifts in organizational structure and processes (perhaps significant shifts depending on how data and data governance is currently being handled). But by tackling the aforementioned challenges from the top down, businesses can ensure that efforts to comply with the GDPR not only go smoothly, but also have larger benefits for the company and the data team. For example, changes made to comply with the GDPR (specifically the changes below and implementation of a data science platform) will likely make teams working with data more efficient overall. And having sound data governance and secure practices also has the added benefit of building customer trust.

Below, we'll walk through each of the immediate challenges and suggested steps for getting started and developing a solution.

1. DATA STORAGE



When it comes to the GDPR, organizations will ultimately need to take stock of where all data is stored and ensure that it is accessible, but only to those with a business need to access it. Data team leaders (and DPOs if they are required for your organization under the GDPR) should be able to easily understand and audit data sources, who has access to what, and what sources are being used for which projects.

To accomplish this, ideally, access to all data (no matter where it's stored) happens from one centralized location. Data science tools/platforms can easily provide this type of governance and connect to multiple data sources, centralizing access to all data (including personal and sensitive data) for the entire organization. To get started, choose a data science platform that has a wide range of data connectors and supports open-source technology to ensure that all current storage systems as well as any future storage innovations are supported.

2. ALIGNING TEAMS

The GDPR changes will certainly force any organization not currently fostering collaboration between teams to do so quickly. But it's not just a matter of increasing communication over email or company chat. There will need to be a certain amount of transparency surrounding data protection that allows a customer service team to field requests without having to ask the data team for an answer every time or the marketing team to understand what the GDPR restrictions are and not inadvertently violate them when completing a customer targeting project. Additionally, data teams working on new projects can communicate back to the legal team responsible for maintenance of the customer consent agreement and can update it accordingly.

Data science platforms can also help facilitate the alignment of teams around a common goal of compliance. When everyone is working with data in the same space, team leaders can keep an eye out, and any violations can immediately be raised and resolved before they expose the company to liability.

3. ACCOMMODATING DATA SUBJECT REQUESTS

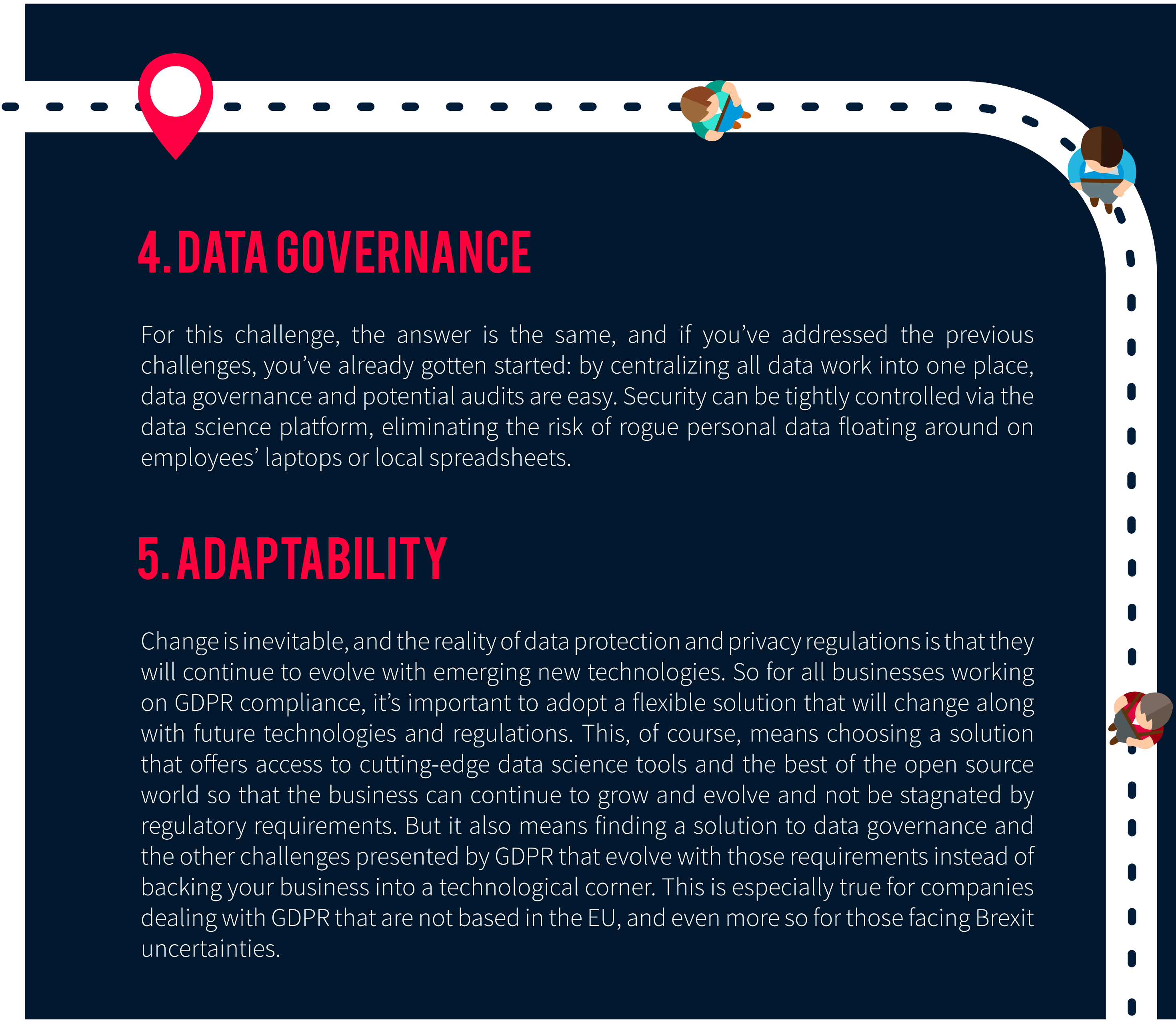
One of the biggest changes with the GDPR is the rights of data subjects. Under the new legislation, data subjects have the right to:

- Be forgotten (have their data erased).
- Access (obtain information about exactly what data is being processed where and for what purpose).
- Data portability (receive a copy of the personal data concerning them).
- Question and fight decisions that affect them that have been made on a purely algorithmic basis.

While it's impossible to predict how many data subject requests you may receive, it's critical to be prepared and have an efficient process in place. And it's not a good idea to wait and develop a process when the first request comes in!

Partially, this challenge hinges heavily on the previous challenge - communication between teams. How will customer-facing teams that are fielding data subject requests satisfy them? **Using a data science platform to have a central place where that team can self-serve and provide the information for the customer without having to ask the data team will likely be the most efficient method.** It has the added benefit of logged access so that in case of questions or an audit, it's very clear who fulfilled these requests and when they were fulfilled.

A data science platform will also provide the necessary transparency around algorithms. Robust data science platforms allow the flexibility of using black box or interpretable models depending on the use case, and for the latter, any requests for justification of decisions decided by an algorithm will be simple to field.



4. DATA GOVERNANCE

For this challenge, the answer is the same, and if you've addressed the previous challenges, you've already gotten started: by centralizing all data work into one place, data governance and potential audits are easy. Security can be tightly controlled via the data science platform, eliminating the risk of rogue personal data floating around on employees' laptops or local spreadsheets.

5. ADAPTABILITY

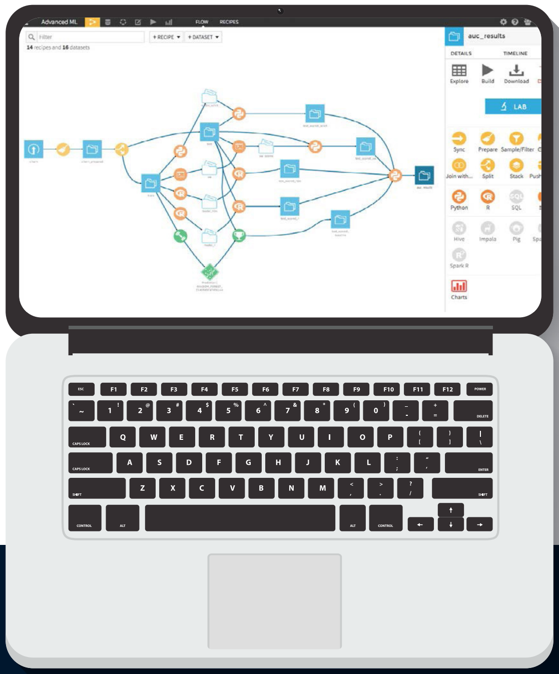
Change is inevitable, and the reality of data protection and privacy regulations is that they will continue to evolve with emerging new technologies. So for all businesses working on GDPR compliance, it's important to adopt a flexible solution that will change along with future technologies and regulations. This, of course, means choosing a solution that offers access to cutting-edge data science tools and the best of the open source world so that the business can continue to grow and evolve and not be stagnated by regulatory requirements. But it also means finding a solution to data governance and the other challenges presented by GDPR that evolve with those requirements instead of backing your business into a technological corner. This is especially true for companies dealing with GDPR that are not based in the EU, and even more so for those facing Brexit uncertainties.

CONCLUSION

These challenges only scratch the surface when it comes to the changes your organization might need to make in order to comply with the new GDPR. But centralizing and standardizing data practices by choosing a data science platform that addresses many components of the regulations at once is a great place to start.

And once these central challenges are resolved, your business will be able to move on to addressing some of the smaller procedural changes and organizational adjustments necessary for full GDPR compliance. You'll likely also find that taking this big first step and investing in a tool for centralized data science and processing addresses many of the other components of the GDPR as well (like the concept of privacy by design). Getting started in GDPR compliance doesn't have to be just something to check off your list. Take advantage of the opportunity to streamline your big data strategy and not only meet regulatory requirements, but empower teams across the organization to become more efficient and scalable.

DATAIKU AND GDPR



About Dataiku Data Science Studio (DSS)

Dataiku develops the unique advanced analytics software solution that enables companies to build and deliver their own data products more efficiently. Its collaborative, team-based user interface works for all profiles, from data scientists to beginner analysts, and the unified framework allows for both development and deployment of data projects.

Features to Support GDPR Compliance

At its core, Dataiku enables organizations of all sizes to employ sound data governance, change management, and monitoring strategies, all of which are the basis of GDPR compliance.

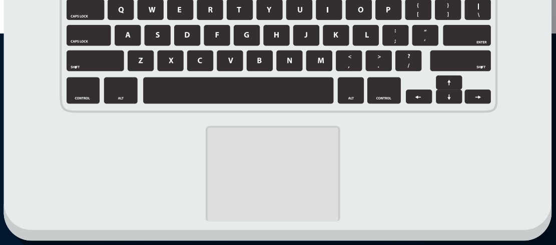
Specifically, our customers are already using Dataiku to:

CENTRALIZE DATA ACCESS

Dataiku allows teams to search for data, comments, features, or models in a centralized catalog. And because all project work is centralized in the tool, team leaders can eliminate the risk of sensitive data on local machines and in spreadsheets - anywhere outside a secure, monitored environment.

ASSIGN ACCESS TO CERTAIN SUBSETS OF DATA

Dataiku allows team leaders to gate access more granularly than at the database level (which is often wider than necessary and exposes the company to risk). Leverage the capability to easily obfuscate, filter data, and assign access to those subsets of data only to certain groups to get started with GDPR compliance.



MONITOR DATA ACCESS

Dataiku not only easily shows who has access to what data in the present moment, but it also exposes the identities of anyone that had access to data in the past. This allows for easy assurance and confirmation that only the right people have access to any given data. Additionally, Dataiku logs all activity to show exactly what specific users have done and with what data. Robust monitoring tools make any auditing or required reporting quick and painless, even under more stringent GDPR restrictions.

TRACK DATA TRANSFORMATION

Dataiku's visual, flow-based interface allows data team leaders (or anyone in charge of data governance or GDPR compliance) to see at a glance what data is being used as a part of which analytics projects and in what algorithms. It shows the entire history of what has been done to data - this makes monitoring data use, and justifying the use of certain data if necessary, seamless.

Use Case

A Dataiku customer in the e-commerce industry had a database containing personal data and several teams working on various projects with that data. To reduce risk and respect customer privacy, they wanted to make sure that people working on a specific project only had access to data relevant to them, not all the data. For example, each marketing team geo was working on their own customer segmentation project where they only needed access to data from customers in their geo.

The global data team leader was able to create a project in Dataiku where she filtered the larger data set into smaller data sets by locale, creating a separate data set for each geo and granting access to the relevant data for each team.

Florian Douetteau

CEO at Dataiku

FEATURED INTERVIEW

Florian is the Chief Executive Officer of Dataiku. Florian started his career at Exalead, an innovative search engine technology company. There, he led an R&D team of 50 brilliant data geeks until the company was bought by Dassault Systemes in 2010. Florian was then CTO at IsCool, a European leader in social gaming, where he managed game analytics and one of the biggest European cloud setups. Florian also served as Lead Data Scientist in various companies such as Criteo, the European Advertising leader.

How does the GDPR affect data ownership?

The main aspect is that a company must apply privacy by design and specify the original purpose for capturing the data, and they must ensure that this purpose is really enforced through the use of that data. So it means companies must have a more proactive approach where they can definitively say "We use this data and this data for this and that purpose." That's really a game changer in terms of how companies will articulate their data strategy in the future.

What are some of the biggest challenges companies face with increased data subject rights (particularly the right to be forgotten)?

The thing is, with GDPR, you already have lots of data that is collected and lots of data that is being used, and all the legacy processes and legacy data that exists in an organization is affected by this change. As soon as GDPR goes into effect, it applies not only going forward, but retroactively.

Today, data collected within an organization is copied many times over, and personal data can be used in lots and lots of different information systems. Let's take a specific example - you can collect a particular piece of information (like a customer's year of birth or the model of car that they own). Maybe you're reusing this particular information over and over, and on top of that, maybe you're sending it to some partners and getting more information back from them about that customer. So in the end, if you have a seemingly simple right to be forgotten request from this user, it may have an impact everywhere throughout your information system, and you might not be able to easily assess where the data is.

And if your system is not built in such a way that you can just remove a piece of information about any customer, this simple request can be a big pain for organizations that are not ready for it wouldn't know how to respond.

So what does best practice look like? What can companies do to start preparing now?

Becoming GDPR compliant is both a matter of preparing your data and also preparing the processes for how you handle, manage and use that data. Companies must be sure that they can provide the full lineage of their data. Lineage means that when you build a particular, let's say, score associated with a customer, you must be able to go back in time and determine which particular data you used to actually build that particular score for this customer.

The reason this is possibly a challenge is because the way information systems are built today is in multiple steps. One group in the organization is building the first step of data, and they give it to the second group in the organization who then refines and processes the data, melds it with a third-party source, and then a third group comes along, and so on and so on. This way, the global lineage of data is something that is very hard to build, and companies need to think about that right now in order to be ready.

In order to work with all this legacy data they have, organizations must start right now to build new systems so that they can trace existing data. They might have to build new systems where they can filter and clean up their existing data in order to be compliant with GDPR. Since it's typically a two year project to get these things done, they must start right now.

What are the most effective strategies companies can adopt to become GDPR compliant?

The most important aspect is to develop a strategy around data projects. Because of the importance of data protection, impact assessments, and the need to specify a purpose and a consent for data collection, you need to set up your data projects. You must know that one particular aspect of your product, one particular asset, or one particular feature is contained within a project. Adopting that idea of learning how to contain those uses of data in separate projects could seem costly, but it's the right thing to do in order to build your architecture in such a way that you can be compliant in the long term.

Companies must also change the way they build their infrastructure and their data projects to stop this step-by-step approach where they've got IT doing part of the job, and some analytic team doing the second part of the job, and a business team doing the third part of the job in a complete separate manner. In terms of specific tooling and approach, most companies will have to start building platforms that handle data end-to-end in a way such that different people in the company can collaborate. Because that's basically the only way you can provide this lineage and structure of data that makes you compliant with GDPR.

KEYRUS
insight into value

Santiago Castro

Head of Strategy and Portfolio at Keyrus

FEATURED INTERVIEW

Santiago has 16 years of experience facilitating change for a wide variety of clients to become data-driven decision making organizations. As a Head of Strategy and Portfolio, Santiago primarily oversees the UK operations for Keyrus UK and also takes on wider Keyrus Group responsibilities, providing deep knowledge and leadership in the areas of reporting, business intelligence, data analytics and decision support systems.

Keyrus creates value in the era of data and digital, helping enterprises take advantage of this paradigm to enhance their performance, assist them with their transformation, and generate new levers for growth and competitiveness.

Which part of GDPR will most affect and impact businesses?

The effects of GDPR will depend heavily on the type of business, and the interesting (or difficult) thing is that there's no one fit-all solution. How GDPR will play out all depends on a business's purpose for holding user or customer information and the type of information. It may be more challenging for certain businesses (for example, financial services organizations) to report, document and govern data because of higher levels of complexity overall.

But in general, what will affect businesses most is not just about documentation and governance but being able to respect certain rights of the citizens or consumers. For example, any customer at any time can ask to be forgotten or to see what's happening with his data - that is what will become more difficult. Because it's not just about reporting what you're doing with data; it's being able to find specific, single pieces of data, attribute it to that particular person who made the request, and take action with it. That's a really granular level, and in that sense, it can be very complex, especially for organizations with lots of customers.

What feedback or response have you had from customers on GDPR so far?

We have three types of customers: 1) the ones that are reactive, 2) the ones that are proactive, and 3) the ones that are unaware. The reactive ones have already had a data breach in the past or have already had an exposure and a business cost (like a loss of reputation, etc.). On the other side of the scale, you have those that are unaware and/or just don't know about GDPR. They either don't know the regulation, don't know they need to plan ahead, or aren't aware of the complexity.

But fortunately as we get closer to the enforcement date, there has been much more awareness. We generally find that as soon as we explain GDPR and what it means, we open up the door to lots of conversations, and they're starting to be proactive and help us help them assess what is needed.

FEATURED INTERVIEW

You've mentioned the regulation is (or can be) very complex; so practically, what are the first steps for businesses?

With our clients, we always start by mapping the data life cycle. If you're B2C, you'll have to look at customer data, or for HR you'll have to look at employee data, so again it depends on your business model. But the general idea is understanding who enters the data, who processes it, who makes copies of data, who erases the data - all the people that touch the data - and we create a map from that. Then looking at that map, add another layer that is: which aspects are related to personal data, which stakeholders (like marketing, finance, etc.) "owns" or has access to or records of those data.

From there, there's another side of the mapping which is looking at the types of customers, partners, and vendors you have (any stakeholders in the data process) as well as current policies and processes. The final step is looking at which parts of the mapping (sources, processes, stakeholders, etc.) represent risk with regards to GDPR compliance, and this makes it easy to find the biggest risks and tackle them..

For some organizations this mapping is big, so they'll have to focus. But at least having the map will allow them to know where to start. and once they know where to start, they can start to tackle from there.

How can companies determine when exactly they need to get started on GDPR compliance?

The output of this mapping is a list of risks. Prioritize that list, and then look - when you think about those risks, where are the gaps? And think… do I need to do a lot of work to close that gap? Evaluate the size of all the gaps in particular, and depending on the issue, you'll be able to develop smaller, more manageable plans of action. Then look, if the first gap will take two months to close, the second two weeks, etc., do you have time to close all the gaps?

Now, some organizations aren't that complex, so they might be able to close all their GDPR gaps in three months. But maybe you have a large organization and realize you need more than a year - that's why we tell everyone not to wait and to do the mapping now. At least then you can discover if you have time to sit back and relax a little or not.

Practically speaking, how can businesses "close the gaps" to become GDPR compliant?

Closing the gaps from the mapping is a combination of things - you'll need to put some process in place, use some technology, and do some training and employee awareness.

That's really important - you may put together lots of technology and process solutions, but if people keep duplicating records and breaking those processes, you aren't progressing and won't be compliant - you're rowing against the current. You need to make people aware of GDPR and its changes (including how it impacts what people are doing) so that when you put a process in place, it works. Train people and give people the education, and let them see how their actions will be affected by the new policies.

Once you deal with the people, then put process in place followed by the tools that allow you to automate the necessary processes brought about by GDPR changes. Tools are definitely needed, but I always like to tell companies to put them in at the end after the people and the processes. If you put the tools first and expect GDPR compliance to happen, you're going to have people who are going against your current.

Any other GDPR advice you give your clients?

Yes - look at GDPR as an opportunity, not just as a cost or a risk or a pain or a panic. It's an opportunity because on one side, what the law wants is for you to demonstrate that you're in control of the data you're collecting, and if you're in control, then you can secure and govern it properly. But this organization and heightened governance comes with the ability to get more value out of your data. If you know what types of data you have and what you're using it for, being more proactive and more innovative, finding more ways to use the data.

On top of that, remember that GDPR was created with the philosophy that customers should be able to trust organizations to hold their data. If you can prove you're in control and that you secure your data, you also can prove that customers can trust you. If customers trust you, they'll be happy to give you data, and you in turn can improve your service.

At the end of all this, you're creating a relationship with customers where there is trust. The more you can prove trustworthy, the more customers will allow you to know about them and the greater number of happy customers you will have. GDPR is not just a risk - it's an opportunity for you to gain market share.

Adrien Basdevant

Attorney at Law, Partner at boutique law firm LYSIAS PARTNERS

LEGAL IN-DEPTH

GDPR: The Law of Large Numbers

When it comes to the General Data Protection Regulation (GDPR), many business are asking: what are the odds of getting caught or paying a large fine? Companies seek to mitigate their exposure by developing a clear understanding of the risk. Entrepreneurs, CEOs, and managers crunch the numbers. It's all about statistics: they want to comply only if the stakes are high.

But a major change is on its way. As the sheer volume of data grows, so do the financial risks associated with data processing. When the GDPR[1] comes into force on May 25th, 2018, data protection laws will be among the strictest legal regulations, along with anti-bribery and anti-trust laws.

Now that we are talking about big numbers and big stakes, people will pay attention. Behaviors will change. GDPR will guarantee long-term results because companies will have weighed their options and opted in favor of compliance over risk of fines.

Will those new regulations stifle innovation and jeopardize competition? This question was the main sticking point of the debate around the GDPR at the European Commission since it was first proposed in January 2012. But that is the wrong way to look at the problem. Companies should instead focus on leveraging the GDPR to their benefit and make its stringent data protection mandate a competitive advantage.

Both businesses that worry about getting caught and those who genuinely want to work to promote data protection should be aware of the what's coming and prepare for it. In particular:

1 Transparency on Data Breaches Will Become the Norm

It took Yahoo almost three years to warn users about what was in 2016 one of the biggest breaches of customer data in history[2].

"In the case of a personal data breach, the controller shall without undue delay and, where feasible, not later than 72 hours after having become aware of it, notify the personal data breach to the supervisory authority competent (…), unless the personal data breach is unlikely to result in a risk to the rights and freedoms of natural persons. Where the notification to the supervisory authority is not made within 72 hours, it shall be accompanied by reasons for the delay" now states Article 33 of the GDPR.

With the datafication of everything, data breaches will become the norm. Even small- and medium-sized businesses (SMB) will be lucrative targets. Prepare for data security to ensure prompt, compliant notification to any breach.

2 Financial Risk Will Rise

It only cost Google €150,000 – the largest-ever fine issued by the French Data Protection Authority (CNIL) – after it had decided to merge all of its existing privacy policies into one single document in 2012 without demonstrating the legal basis for such combination[3].

Google Street View, despite holding the record for unlawful collection of personal data on a massive scale, only cost the company €100,000 in France (March 2011) and €900,000 in Spain (December 2013). That represents just about 0.001% of Google's annual turnover, or less than six minutes' worth of business for the company[4].

"Infringements of the following provisions shall (…) be subject to administrative fines up to 20,000,000 EUR, or in the case of an undertaking, up to 4 % of the total worldwide annual turnover of the preceding financial year, whichever is higher" provides Article 83-5 of the GDPR.

GDPR introduces drastically higher sanctions for non-compliance that will be split into two broad categories.

A first category of fines up to €10,000,000 or up to 2 percent of total worldwide turnover of the preceding year (Art. 83-4) applies, among others, to breaches of:

- Consent to the processing of data relating to children (Art. 8)
- Data protection by design and default (Art. 25)
- Maintain written records (Art. 30)
- Implement technical and organizational measures (Art. 32)
- Report data breaches (Art. 33, 34)
- Conduct of privacy impact assessment (Art. 35, 36)
- Appoint Data Protection Officers (Art. 37 to 39)

The higher categories of up to € 20,000,000 or 4 percent of total worldwide turnover of the preceding year (Art. 83-5) will apply specifically to breaches of:

- The basic principles for processing, including conditions for consent (Art. 5, 6, 7 and 9)
- Data subjects' rights (Art. 12 to 22)
- International transfers (Art. 44 to 49)
- Non-compliance with an order imposed by supervisory authorities (Art. 58-2)

This clearly shows GDPR is more than just a legal challenge. It invites radical management and cultural change, and it forces every stakeholder to transform the way they collect, combine, process, share, and use data.

3 Data Will Be the Future of Litigation

Keep in mind that being a data controller or data processor does not change the fact that supervisory authorities are empowered to impose significant administrative fines.

This regulation harmonizes the protection of fundamental rights and freedoms of natural persons with respect to processing activities among all EU Member States. Data will be at the heart of every type of litigation: from hacking to discrimination in algorithmic consumer profiling.

4 Algorithmic Accountability Will Come to the Forefront

Businesses collect more and more information from their users, but users don't know how companies make decisions about them in relation in the context of online advertising, hiring, lending, etc. So civil society and privacy advocates are led to believe that algorithms act invisibly, without accountability, and sometimes without citizens even knowing they are being profiled and targeted.

Should companies disclose how algorithms use personal data? Or should we regard those algorithms as intellectual property; trade secrets that ought to remain intrinsically mysterious? Companies should take a strong stance on this ongoing debate to differentiate from competitors.

Companies will soon have to integrate these questions in their strategy. Privacy and algorithmic accountability could be poised to become a competitive advantage or a core element of the customer relationship.

[1] REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)
[2] Yahoo Says 1 Billion User Accounts Were Hacked, The New York Times, December 2016
[3] CNIL, Deliberation n°2013-420, January 3rd, 2014
[4] Google's revenue worldwide from 2002 to 2016 (in billion U.S. dollars)

GoDataDriven
proudly part of Xebia Group

Giovanni Lanzani

Chief Data Science Officer at GoDataDriven

Giovanni holds a PhD in Theoretical Physics, with research focusing on models describing DNA mechanics. Prior to joining GoDataDriven, Giovanni worked at the Software Quality department at KPMG.

As Chief Science Officer Giovanni advices clients on how to extract valuable insights from (big) data analysis.

THE IMPACT OF GDPR ON ALGORITHMS

GoDataDriven develops artificial intelligence solutions for data driven organizations, and organizations that aspire to become one. The team consists of a unique combination of both Data Engineers and Data Scientists, that combine their academic background with state-of-the-art know-how of technology. Experimentation and innovation are essential elements of GoDataDriven's approach, that is always focused on developing scalable, well-structured, solutions that are taken into production.

GoDataDriven recently launched the Big Data Survey to discover how organizations really use big data, and one of the questions was about the respondent's familiarity with the EU General Data Protection Regulation (GDPR).

The regulation will be enforced starting May 25, 2018, and the required changes are (especially for some organizations) far reaching. It was therefore a surprise to learn that a third[1] of the Survey respondents are not familiar with the regulation.

Specifications of GDPR

The GDPR is far-reaching, but to recap, some of its most important provisions specify that:

- Automated decision making is contestable, meaning the outcomes of algorithms are questionable and fightable.
- Automated decision making should not use personal characteristics like age, race, etc.
- Data protection must be implemented by design and by default.
- A Data Protection Officer (DPO) must be appointed for certain companies.
- Revoking consent for using data for a specific purpose must be as easy as giving consent, and that consent can be revoked at any time.
- Pseudonymisation should happen as soon as possible.
- There is a right to erasure.
- Data should be portable between data controllers.

Are Organizations Prepared?

In any large-enough organization, things such as right to erasure, contestable decision making, and revoking consent are changes that sweep across many systems and units. Individually they represent a lot of work, and together they almost seem daunting.

[1] 100 out of 276 respondents.

THE IMPACT OF GDPR ON ALGORITHMS

Many companies are struggling with the GDPR already. Almost half of survey respondents[2] said that their company is not ready for the GDPR. The other half presumably has not realized what the GDPR really means for their company!

Not surprisingly, there are a plethora of articles (just do a Google search for "how does the GDPR affect me") that explain what the GDPR is, how it affects companies, and how to adapt. But most of them present the GDPR from the risk angle: what should you do to be compliant with the GDPR to avoid the steep fines[3] of non-compliance?

What we don't see much is the opportunity angle, which is much more exciting (and less scary).

The Other Side of GDPR

The opportunity angle arises from the data portability section of the GDPR (Article 20, Right to Data Portability). Paraphrased, it states that data subjects have the right to receive personal data concerning them and that they provided to the data controller.

The caveat is that the data cannot be dumped on the requester in a raw state (as it happened in [Europe vs Facebook][4]): it must come in a "commonly used and machine-readable format." Moreover, when technically feasible, "the data subject shall have the right to have the personal data transmitted directly from one controller to another."

If you think about it for a second, you can immediately see that when your algorithms get better with more data, your company becomes more attractive for your potential clients.

For example, let's say you are an insurance company collecting driving behavior and adjusting the premiums using that data. The more years of data you have, the greater the potential savings for your customers. If you are the first (or the only) provider accepting data from others, GDPR poses a great opportunity for your organization.

What you need is to be quick in creating converters from the format your competitor is using and the format you're using. With that, your onboarding becomes much smoother, and you can immediately use this to your advantage. On the other hand, this also means that the idea of customers being locked in to your service doesn't hold anymore, and you'll have to get creative and differentiate along other axes.

Take the example of an online shop. If customers can bring their data with them, that means you can offer new services where fashion recommendations are much more tailored to customers' real style[5], as your algorithms will have a much richer history at their disposal.

There are many more examples, and I'm sure each company has its own use case where getting data from competitors would be of value. So the next time you feel your head spinning because of all the new burdens that the GDPR might bring upon your company, don't forget that there are plenty of opportunities made possible by GDPR as well.

[2] Of the ones that didn't reply with "I don't know."

[3] Up to €20M or 4 percent of annual worldwide turnover, whichever is greater.

[4] http://www.wired.co.uk/article/privacy-versus-facebook

[5] This of course assumes that it will be possible and seamless to share your shopping patterns.