White Paper

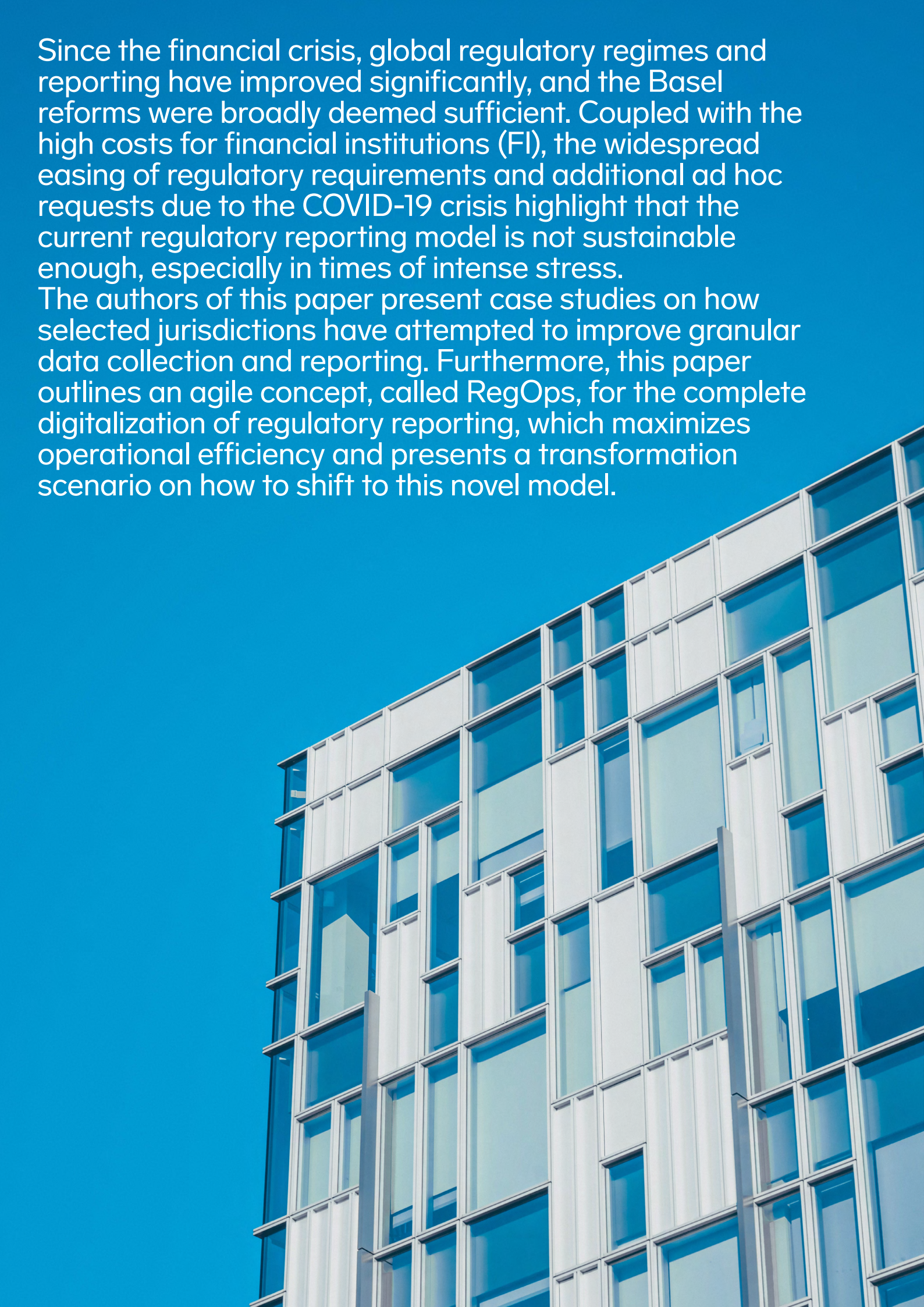
Martina Drvar, Vice governor Banking Supervision, Croatian National Bank
Dr. Johannes Turner, Director Statistics Department, Österreichische Nationalbank
Dr. Maciej Piechocki, Member RegTech Management Board, Regnology
Eric Stiegeler, Senior Manager, Regnology
Daniel Münch, Business Advisor, Regnology

RegOps

The future of data collection and data management: Agile RegOps for digitalizing the regulatory value chain

REGNOLOGY

Since the financial crisis, global regulatory regimes and reporting have improved significantly, and the Basel reforms were broadly deemed sufficient. Coupled with the high costs for financial institutions (FI), the widespread easing of regulatory requirements and additional ad hoc requests due to the COVID-19 crisis highlight that the current regulatory reporting model is not sustainable enough, especially in times of intense stress.
The authors of this paper present case studies on how selected jurisdictions have attempted to improve granular data collection and reporting. Furthermore, this paper outlines an agile concept, called RegOps, for the complete digitalization of regulatory reporting, which maximizes operational efficiency and presents a transformation scenario on how to shift to this novel model.



Introduction

The decade of the 2010s saw the implementation of Basel III reforms (BIS 2011) to cope with fallout of the 2007/2008 financial crisis. With the official end of the post-2007 crisis agenda (BIS 2020) and the Covid-19 pandemic as well as the ensuing economic crisis, we are at a perfect point to evaluate banking regulation and the corresponding banking regulatory reporting system.

When looking at the results of the Basel III reform (BIS 2019) regarding its performance, one can observe a mixed picture. For one, many of the ideas manifested in the Basel reforms and its national and supranational implementations were highly successful. Especially higher capital requirements and increased resilience of financial markets have proven to be a stabilizing factor for the world's economies instead of financial institutions being the primary source of instability. Overall, the reforms had a positive influence on the well-being of the world economy.

On the other hand, several major deficiencies of banking regulation became visible or are now prominently in the focus of all involved stakeholders in the financial market. Most remaining shortcomings are interestingly found not in the principal ideas and concepts of the regulatory reforms but their functional, technical, and organizational implementation. To find approaches to overcome these shortcomings, the authors propose the "RegOps" approach, a regulatory reporting framework that combines an integrated data flow, a common processing of standardized, granular datasets based on a big data-enabled platform for computation and analysis. This model was also successfully implemented in a Proof-of-Concept and proved to deliver all requirements conceptually.

The authors would like to embark on this journey by citing Benoît Cœuré (2020), Head of the BIS Innovation Hub: "The benefits and opportunities of regtech and suptech for regulated entities and supervisory authorities to improve efficiency, reduce manual processes and make effective use of data are enormous. As they are more widely adopted, these technologies can enhance diligence and vigilance in risk monitoring and management in real time, improving the resilience and stability of the broader financial system.

Problems

The first major deficiency is the generally low application of innovative technology in the fields of digitization and modern computing in the banking regulatory reporting regime. In most regulatory frameworks in global jurisdictions, regulatory data flow still happens in a quasi-manual, template-based fashion. This means that the mere automatization of manual, printed, or handwritten reporting processes of aggregated data, which was the main activity in the past years, is not enough.

The digitalization of regulatory reporting does not only mean changing technology but also requires rethinking the whole process, from the beginning of data generation within banks throughout the entire processing chain to the regulators and analysts. Only a few jurisdictions have started the journey thinking regulation anew and leveraging the possibilities of new technologies like artificial intelligence (AI), application programming interfaces (API), big data, the cloud, high-performance computing, and blockchain/distributed ledger technologies (DLT). Many of these technologies, which were just mere buzzwords a few years ago, have meanwhile matured far enough to contribute to and enable new approaches to banking regulation, and especially regulatory reporting. What is necessary for regulators is to understand how to build a functioning architecture by combining innovative technology and transforming existing frameworks in a future-ready state.

These technologies could also help to topple the second point, the high cost of regulation and regulatory reporting. Estimations of the cost of regulatory reporting vary wildly, yet all indicate very high costs for financial institutions. McKinsey estimates that the annual cost for regulatory reporting of UK banks is 2bn – 4.5bn GBP (Van Steenis, 2019). A commission staff working document estimated 4bn EUR for the European Union (European Commission, 2020), while a study by Chartis & BearingPoint estimated the cost of compliance in the EU and the USA for the full scope of risk data aggregation and regulatory reporting to be approximately 70bn USD (Chartis and BearingPoint, 2018).

While banking regulation has become more effective over the last decade, it is clear that the marginal use of an ever-increasing set of template-based regulatory requirements is strongly decreasing; the main impediments being the limited insight and flexibility of the aggregated data reported. Also, it becomes clear that while technology could significantly reduce costs, it currently cannot be deployed efficiently because of a lack of common standards in data models and processing. The financial markets would need a common standard to describe regulatory data requirements and the corresponding regulatory logic processing before leveraging large amounts of data with modern technology. To a large extent, the current high costs in regulatory data generation for institutions are rooted in the necessity to leverage the same information artefacts over and over again for different non-aligned regulatory reporting regimes with myriads of templates (prudential, national, statistical, granular, resolution reporting) with often very similar, but slightly differing definitions. Institutions' costs could be sharply decreased if data would only be requested once in a granular, standardized, aligned, structured fashion, and processed with common regulatory logic while supervisors would have better data quality, much more agile access to data and a far increased flexibility to get the answers to the questions they are truly interested in. Improved, common forensic insight in risk concentration would optimally lead to reduced losses for banks and more insight for regulators concurrently. Standardized granular data models and processing logic would also boost the use of new technology and strongly decrease implementation costs and costs of change, which are currently one of the main hurdles of technology adoption. However, common standards require finding governance models between the different stakeholders within the financial markets on the one hand, and between the financial market stakeholders and the regulatory authorities, on the other hand.

The third shortcoming of the current regulatory regimes is the lack of operational excellence, which became apparent via several high-profile failures in recent years, such as the hidden derivative losses at Banca Monte dei Paschi di Siena (Sanderson and Crow, 2019) and the Wirecard scandal (McCrum, 2020). For one, offsite supervisory overview is still limited due to the nature of the collected data. Aggregated and template-based reporting is conceptually more prone to data correction or even manipulation. Fully granular, automatically pushed, end-to-end integrated data delivery, possibly accompanied by other trust ensuring technologies like blockchain, could strongly improve trust and operational stability for data reporting and could data manipulation virtually impossible or prohibitively expensive. Another problem is the lack of quality, timeliness, and inter-entity matching, meaning the complementary fit of the two datasets, representing two sides of the same transaction, which could be strongly improved using granular, end-to-end delivery of granular data.

On the other hand, the Covid-19 crisis as the first real test of the new regulatory regime showed that while many of the Basel reforms, in general, had a positive impact, the regulatory reporting component proved difficult to operate. The paradox situation arose that many jurisdictions issued moratoria on new reporting regulations or eased reporting obligations exactly during the crisis where ample information is critical for regulators to make informed decisions. Even if eased obligations primarily concerned less important information for crisis management, we could see that the implementation of urgently needed new data requirements lasts many months (e.g., data on moratoria and state guarantees), which is too slow.

Stock-taking exercise & case studies

The following part aims to outline a few main elements needed to improve the current regulatory reporting regimes in the authors' view. A stock-taking exercise will follow to see how certain notable jurisdictions found solutions to address these needed elements, and why they were successful to introduce them. The following features are missing in current regulatory reporting regimes:

Standardized input data model	Regular	
	Market	
Pull-mechanism		
Integrated data transfer		
Granular data delivery		
Big data-enabled analysis tool		

- **Standardized input data model and processing logic:** standard of data and data processing used for regulatory reporting, which is either defined by the regulator or the market participants
- **Pull mechanism:** the bank does not submit data to the regulator (push of data); instead, the regulator accesses the required data (pull of data)
- **Integrated data transfer:** end-to-end data flow, which is fully integrated and automatable via modern interfacing, e.g., API
- **Granular data delivery:** banks submit contract-granular data instead of aggregated regulatory reporting templates to fulfill the regulatory requirements
- **Big data-enabled platform:** employing a big data-capable regulatory platform which can collect, store, and analyze large amounts of granular data to generate insights for authorities

Traditional and current market approaches

The first regulatory reporting requirements in the 1970s and 1980s were processed manually, mainly filing paper templates with a pencil or pen. Since the inception of IT-based regulatory reporting, solutions have sought to answer the fundamental question of how data is optimally mapped from the source systems onto the reporting templates. Starting from internally developed solutions, it became increasingly clear that there may be efficiency gains using standard software from external solution providers. These solutions providers have used three main approaches to solve the data sourcing issue:

a) Last mile/typewriter approach
This solution offers only an electronic version of the reporting template, and the user types in the cell values and contents directly into the templates. All calculations, transformations, and consistency checks of input data have to be done manually or via other individual tools, thus eliminating the need to build a working data connection and pre-processing from the source system. Admittedly, this system became increasingly sophisticated by introducing innovations like electronic reporting files and template-based electronic validation, e.,g. the European Banking Authority's (EBA) XBRL DPM (Data Point Model) initiative between 2008 and 2013. Still, essentially the solution is only a digitalized version of aggregated, paper-based templates. While being easily deployable, integrable, and usable, this approach clearly lacks the sophistication needed for more complex, automated setups.

b) Institution-individual data sourcing & modeling
In this approach, the standard solution provider makes available a general toolset for mapping data from data sources, building transformation, mapping, consistency checking rules, displaying templates, and generating template-based reporting files. This option is highly flexible and offers financial institutions a high degree of freedom in customizing the components to their needs.

However, this approach requires a large IT and functional staff to integrate, customize, maintain, and operate the system. It often requires separate setups and IT builds to cover different reporting requirements. Furthermore, an institution following this approach needs to find functional and IT-architectural answers to regulatory requirements independently.

c) Standardized data sourcing & modeling
In this approach, the standard solution provider requires the financial institution to deliver a multitude of data requirements formatted in a standardized, granular input data model and processed by common logic, covering the requirements of multiple reporting frameworks. This approach, if done correctly, promotes reconciliation across reporting by design and non-redundancy of information. The mapping of the institution-individual data sources to the standardized input data model needs to be mostly customized. It requires a relatively high one-off investment for building the interfaces. However, after the initial mapping exercise, the solution provider can apply different transformations, mappings, and consistency checks upon this database to fulfill various regulatory data requirements with only one data delivery set in the common data model. The reuse of processing logic and the data model at different customers offers the potential of substantial cost reductions via economies of scale, i.e., positive effects on operational, IT sourcing, and IT infrastructure cost.

Furthermore, the model encourages financial institutions to work together on regulatory topics and align a common understanding vis-à-vis the regulatory authorities, streamlining communication, and enabling a mutual understanding of data flow and calculation. There are attempts of regulatory authorities to foster bespoke data models and processing logic, i.e., the Banks' Integrated Reporting Dictionary (BIRD) (European Central Bank, 2020) or the Integrated Reporting Framework (IReF) (European Central Bank, 2019). However, currently, only market solutions offered by commercial vendors are productively in use and possess the maturity to be deployed.

All these approaches can make sense in individual use cases and are often used in parallel by different entities of a financial group or even within individual entities within the same financial institution. Yet approach c) "standardized data sourcing & modeling" has emerged as the most promising model for the future and is applied or targeted for future use by most financial institutions. This model also gives a foundation for the full range of benefits from applying modern IT technology like cloud computing. Many of the leading banking regulatory software providers are already following this approach or have begun the journey of migrating to it.

Switzerland/Liechtenstein (Standardization of source systems)

Standardized input data model	Regular	
	Market	X
Pull-mechanism		(X)
Integrated data transfer		
Granular data delivery		
Big data-enabled analysis tool		

In Switzerland and Liechtenstein, the market for core banking systems is highly consolidated, and most of the Swiss banks are using one of the popular six core banking systems. This consolidation reduced the typical heterogeneity of banking systems as source systems of data. It facilitated the use of standardized data models by implementing six standard interfaces to each of the six core banking systems. It thereby follows the "standardized data sourcing & modeling" approach in section 3.1 and leverages the consolidated market of the banking system, which allows for implementing the interface only six times to cover the majority of the banks.

This standardized data model was developed by market participants and evolved, triggered by regulatory changes. Advantages are reduced costs and more efficient regulatory change management. These banking systems contain a ready-to-use interface that maps the banking system's operational data to the standardized data model used for regulatory reporting. Therefore, new banks joining the market solution can immediately use the interface and save costs for implementation projects and for maintaining the interfaces shared across other banks using the same system. It also enables faster implementation of changes in the regulation. Another advantage of the standardized data model is that it creates a common language to communicate regulatory data between banks and software providers.

As an indirect effect, the standard data model through standardized interfaces also ensures better quality reporting for the regulator because of the shared knowledge about the interface between the banks and a common data dictionary. Thanks to regular exchanges between the regulator and the involved stakeholders, the regulator is often aware of available data in the standard data model and takes the available data into account when drafting new regulations or new reporting requirements.

The reduced number of banking systems on the data input side and interfaces to a standardized data model from these banking systems enable economies of scale for the costs of interpretation, mapping of operational data, and change management of regulatory requirements.

Oesterreichische Nationalbank (AuRep: a standardization of regulatory data model)

Standardized input data model	Regular	
	Market	X
Pull-mechanism		(X)
Integrated data transfer		(X)
Granular data delivery		X
Big data-enabled analysis tool		

In Austria in 2014, the Oesterreichische Nationalbank (Austrian National Bank, OeNB) took the initiative to modernize its regulatory reporting with the Austrian Reporting Services GmbH (AuRep) model (Kienecker, Sedlacek & Turner, 2018).

The first element is the introduction of a common input data dictionary, the Basic Cube and an obligatory output reporting data model , the Smart Cube, which has been developed and is maintained by the OeNB. The Smart Cube was finalized for statistical and financial market stability purposes due to the maximum harmonization requirements covering supervisory statistics only on the input side. The second key innovation from a regulatory standpoint has been introducing granular data delivery of multidimensional sets via this Smart Cube model. The third innovation has been the joint transformation of the delivered input Basic Cube data into the Smart Cube result data and other template-based reporting requirements (especially for supervisory statistics) via one common, standardized logic on the IT architecture of the AuRep utility for almost the entire Austrian banking sector. Although the Basic Cube is in principle voluntary, it has been implemented – mostly 1:1 – by nearly all Austrian banks.

Based on these three elements, the financial institutions ultimately only deliver a granular Smart Cube data set to the regulator (for statistical and financial market stability purposes). These Smart Cube data sets allow the OeNB to receive granular, standardized data sets, which can then be flexibly employed for standard reporting and ad-hoc reporting based on the attributes included in the Smart Cubes. In case the OeNB needs a different view on the data, it can efficiently generate new reporting formats with its regulatory toolset without any implementation cost.

The AuRep model lowered change and implementation costs for the financial market while increasing data quality and traceability for the OeNB. It furthermore enabled institutions to use standardized IT solutions for the entire market. The scope has already been enlarged to the European AnaCredit and SHSG framework and some other national and European supervision and resolution reporting requirements. It is currently planned to onboard other European reporting requirements like CoRep.

Croatian National Bank (CNB BIRD)

Standardized input data model	Regular	X
	Market	
Pull-mechanism		
Integrated data transfer		
Granular data delivery		X
Big data-enabled analysis tool		

The Croatian National Bank (CNB) started a project to define a granular data collection "CNB Banks' Integrated Reporting Dictionary" (CNB BIRD) as preparation for an expected increase of supervisory requirements because of the upcoming EU membership. The CNB BIRD model is based on granular input data and is used for statistical and regulatory purposes. It was driven by the regulator in collaboration with banks and includes about 40 attributes and 19 record types (Bašić, 2017).

In this model, the CNB collects contract-granular data for legal entities and semi-aggregated data for households. To fulfill the regulatory requirements, banks submit their granular data to the CNB, where it gets processed, and the CNB generates the reports. For this purpose the CNB maintains an open, common code repository that contains all transformation logic and allocation rules for reporting templates used by the CNB to process the data.

The banks are then supplied with the results and check them for correctness, and if needed, interfere. This process enables the CNB to collect highly informative, high-quality data for the financial market and ultimately allows them to make better-informed decisions thanks to the increased transparency of the granular data at hand and the ability to drill-down into the data using business intelligence software (BI) and dashboards.

Bangko Sentral Ng Pilipinas (API-based Prudential Reporting System)

Standardized input data model	Regular	
	Market	
Pull-mechanism		X
Integrated data transfer		X
Granular data delivery		
Big data-enabled analysis tool		X

The central bank of the Republic of the Philippines, the Bangko Sentral ng Pilipinas (BSP), and the RegTech for Regulators Accelerator R²A together developed a prototype for an API-based Prudential Reporting System. In such a system, banks will individually prepare template-based reporting in a common data point model (DPM) to BSP in XML format that will automatically be extracted from the supervised financial institutions' databases via an Application Programming Interface (API).

The API establishes a direct link between the supervised banks' pre-processed regulatory reporting data and the BSP. In doing so, pre-calculated aggregate template data (approx. 50,000 data points) is extracted via DPM from banks' databases and forwarded directly to BSP's processing and validation queue in a single reporting package. After processing, the data is passed automatically into a centralized and secure database for storage and more efficient database management. A centralized database can then facilitate analytical tools, such as big data, machine learning, and artificial intelligence.

This API-based approach has several benefits over the previous email-based reporting system, which requires the submission of information in the form of excel tables. It automates the collection, processing, and analysis of template data of supervised financial institutions. Moreover, it could deliver a greater volume of data at faster intervals, with fewer duplications, errors, and omissions. By streamlining reporting requirements, the number of data points was decreased from approximately 107,000 to 50,000 because duplicated information was eliminated (di Castri et al., 2020).

Overall, the prototype demonstrates a streamlined data transmission process of regulatory reporting, including transmission, processing, validation, warehousing, and analysis of financial institutions' prudential reports. Such an API-based prudential reporting system can improve the quality and timeliness of the collection of supervisory data.

Proposed approach "RegOps"

In the previous section, it has been shown that there are several innovative concepts to regulatory reporting in production, which partly deliver the necessary features for future-proof regulatory reporting. However, it can be concluded that none of these innovative approaches fully cover all the mentioned features yet. To combine all these features, the authors propose the RegOps model for regulatory reporting.

RegOps is closely connected to the term DevOps (a portmanteau of development and operation), known from software development and seen as the answer to the shortcomings of the waterfall model. The waterfall model, as a traditional plan-driven approach to software development, has been around for decades. Critics argue that the waterfall model lacks the flexibility to accommodate customer changes and that its linear stages to software development are not people-centered, do not encourage customer collaboration and leave no room for creativity nor innovation. To improve software development, individuals have adopted methodologies that focus on customer collaboration, continuous delivery, constant feedback and communication between developers, customers, and users while delivering software incrementally in small releases. These methodologies have led many individuals to become advocates of an agile way of thinking. Gartner declared that "DevOps movement was born of the need to improve the agility of IT service delivery and emphasizes people and culture and seeks to improve collaboration between development and operations teams while seeking to remove the unnecessary impediments to service and application delivery by making use of agile and lean concepts." (Wurster et al., 2013)

For this reason, DevOps can be considered as the integration and application of different software development methodologies, operational processes, and social psychological beliefs for transforming IT service delivery. From this perspective, DevOps is a new way of thinking, a spirit, a philosophy for transforming organizations. Gartner analysts declare that DevOps "... is a culture shift designed to improve quality of solutions that are business-oriented and rapidly evolving and can be easily molded to today's needs." (Wurster et al., 2013)

Regulation has been developed (conceptualized, drafted, released) and rolled out according to the waterfall model over the decades, leading to disastrous, purely reactive time-to-market and offering hardly any flexibility in embracing regulatory change. Most of all, it created enormous costs to regulators but, more importantly, to the financial services industry. Similar to DevOps, RegOps improves the way regulators and regulated entities interact: collaboration, continuous delivery, constant feedback and communication between regulators and the regulated, while delivering regulatory change incrementally in small releases without affecting the whole system.

RegOps is defined as an approach to systematically change how regulation is developed and deployed and how data is exchanged between regulators and regulated using push and pull approaches. With standardization and industrialization, RegOps provides a framework and infrastructure to regulators worldwide to collect data efficiently and flexibly from the regulated markets.

With the use of modern technology and proven standardization artefacts, RegOps allows regulators to arrive closer to the dream of RegTech from Andy Haldane (Bank of England) in 2014: "I have a dream. It is futuristic, but realistic. It involves a Star Trek chair and a bank of monitors. It would involve tracking the global flow of funds in close to real time (from a Star Trek chair using a bank of monitors), in much the same way as happens with global weather systems and global internet traffic. Its centerpiece would be a global map of financial flows, charting spill-overs and correlations." (Haldane, 2014). With RegOps, regulation and reporting are not a top-down process based on macroeconomic risk considerations, which are transformed into standardized regulatory approaches, definitions, and later implemented in fixed, low insight-giving regulatory templates. Instead, the proposition is to start regulation as a bottom-up process focusing on regulatory micro definitions of standardized data fields and models on the granular dataset and data information level, which can then be flexibly used for any macro-regulatory requirement, such as calculations for different and even changing macro regulatory approaches based on the granular standardized data.

In this section a proposal for the future of regulatory reporting is outlined which not only comprises all the features mentioned earlier, but which could also be deployed in reasonable timeframe due to the availability of its three basic elements. These elements are:

- A unified, normalized, universal data model and standardized, common regulatory logic for prudential, statistical, financial, and resolution regulatory reporting purposes
- A fully integrated, bi-directional data delivery stream including a toolset to export, transform, and load data to deliver functionally valid results
- A big data-enabled platform to collect, store, and analyze data to gain insights for regulators

In the next paragraphs, each of these components will be discussed in more detail.

Standardized data model & processing logic

Data model

A common standardized input data model (IDD) capable of capturing contract-granular financial transactions information represents the starting point and the foundation of the data flow. It acts as a common language for all regulatory reporting and in essence defines the scope, abilities, and limitations of what can be done within the regulatory process. Such a data model needs to be able to capture a full picture of financial institutions of all sizes and business models, and it also needs to cover all financial products. The IDD would need to source data information covering the whole institution and act as a single source of truth for all regulatory information. The principle in selecting the number of artefacts to be captured should be "as granular as necessary, yet as simple as possible."
One challenge is to find common ground on a global scale to define such a standard because of the heterogeneity of regulatory definitions of the Basel Framework (BIS 2020b) and differing reporting requirements per country. Another challenge is adopting such a standard because of the wide range of different data source systems (banking systems) and the lack of common interfaces and shared data formats. While seemingly trivial, the greenfield creation of a productively usable regulatory data model covering all regulatory needs is very resource and time-intensive and most likely not an efficient or sustainable model.

In principle, there are two possibilities to find a common approach for these data models. One approach would be a data model published and maintained by the regulatory authorities ("Active data models" for regulators). Another option would be using available market standards for regulatory data models provided by third-party providers and already in productive use ("Passive data models" for regulators). In a real-world implementation, it would, of course, be most efficient to conceptualize a hybrid model making use of existing standards as much as possible and adapt them to the actual regulatory use cases as needed by the authorities. The current standardized data models developed by the market are working successfully but are neither implemented at all banks within one jurisdiction nor globally. On the other hand, if the regulator designs such a standardized data model in isolation, it might suffer design flaws because of the lack of practical experience and insights on working and design in the data model based on banks' operational data. Therefore, we propose a collaboration between market solutions providers and regulators, focusing on basic data that offers the most added value and the easiest implementation, enables fast adoption for rapidly taking the first steps.

Therefore, a joint, holistic review of the regulatory definitions done by all stakeholders together would allow to unify and consolidate these definitions as a regulatory glossary and could act as the founding pillar of such a standard. This initial cooperation should be leveraged as the nucleus of a joint governing model. There needs to be a mechanism defined to find rational consensus between regulators and banks, but that also guarantees the fast delivery of results of IDD progression and development in a reasonable timeframe.

For optimized data volume, such a data model would typically be designed as a normalized, relational data model with corresponding metadata at least. Besides the technical challenges, the governance of such an IDD needs to fit the definitions, adoption, and maintenance of the IDD. Currently, a globally applicable IDD would require around 1,500 – 2,000 data fields to comply with all current regulatory regimes.

A fully granular, stable data model that clearly defines the scope of the collected information would ultimately free financial institutions and regulatory authorities from the burden of a multitude of regulatory standards by breaking down the required information into the most atomic elements possible. According to various regulatory and accounting standards, information artefacts can then be flexibly aggregated or calculated based on this atomic information.

Excursion: A lean IDD ("less than 100 data fields standardized input model")?

Defining a standard input data model on a contract-granular level promises a high degree of flexibility to use regulatory data for a wide range of regulatory requirements. But on the other hand, a granular data level comes with a wide range of possible data fields, which could be included and might reduce the speed of adoption and increase the cost of maintenance. To enable the fast adoption of such an IDD, it must be as granular as necessary and as simple as possible.

Looking at regulatory reports, it is surprising that available international identifiers are still not used, or identifiers are being requested together with redundant information, which could be easily derived. Looking at the "Legal Entity Identifier" (LEI) code as a unique identifier of entities itself already provides a rich source of information about the entity, the group structure (which would help to identify intercompany entities and positions as well), and the domicile. Other standardized identifiers, e.g., a "products code" like the "Financial Instrument Global Identifier" (FIGI) (OpenFIGI, 2020), "unique product identifier" (UPI) (BIS & IOSCO, 2017), and the "International Securities Identification Number" (ISIN) are needed to reduce data fields in such an IDD to describe recurring templates of products and required data fields.

Regulators could also provide logic in the form of code to reduce ambiguity for mapping to such an IDD. One such example would be an industry taxonomy like the "nomenclature statistique des activités économiques dans la Communauté européenne" (NACE), which a regulator could use to provide a precise mapping of industry taxonomy to counterparty definitions and risk weights used in the credit risk.

A regulatory data model with 100 data fields seems ambitious when looking at different regulatory reporting requirements currently in place across jurisdictions. A switch to an IDD provides a new perspective based on more insights thanks to the nature of the granular data. It triggers the question if the current regulatory data and data information requirements are still appropriate for such a changed environment. In this sense, regulatory data requirements are not static. Thus a model that delivers a limited set of data fields in a much more granular, timely, high quality, and standardized fashion could be a superior supervisory tool to a legacy system that captures a wide array of complicated, non-standardized data fields. The added value of granular data might make current regulatory approaches redundant, provide only additional marginal information, or ask for new regulatory approaches leveraging this granular data. A completely new approach would also offer the possibility to start data collection anew with a lean IDD and provide flexibility to explore which regulatory data field requirements are needed and useful to fulfill the mission of the regulatory agencies. Looking at certain jurisdictions, we can see that lean granular data models, such as in Croatia (~40 data fields) and Switzerland (~80 data fields), are challenging but possible. However, a more comprehensive approach would need a global and supranational alignment on rethinking standards of banking regulation and regulatory reporting.

The implications of such a granular IDD would also mean thinking less about possible approaches and more about data itself (bottom-up approach). Databases on the regulators' side would, in this proposition, replace templates and reports. Calculations like risk-weighted assets (RWA) could be derived or validated at the granular level.

Having a standardized input in in form of an IDD would already be a massive step towards better and more efficient processes by having clear definitions of non-redundant data fields, presenting a single point of truth. Regulatory changes requests would look into the available data source to first (bottom-up) reconcile which data is available and make data-driven decisions to increase the likelihood that a regulatory change will use existing data. This would vastly reduce change requirements on the data input side.

Standardized processing logic

Based on a common language for the whole regulatory value chain, established in an IDD, all regulatory transformations and allocations can be done with a fully standardized, machine-executable regulatory logic in code.

The transformation includes all calculations and derivations processed on the input data delivered by the institutions. This could be the derivation of information needed for specific reporting regimes (e.g., mapping a general industry code to a reporting regime-specific industry classification). But it can also be the calculation of more elaborate key values (e.g., calculating an exposure at default or risk-weighted assets based on delivered inputs for a standardized credit risk approach).

Based on the outputs of the transformation, it is possible to apply an aggregation logic to generate the legacy reporting templates. While standardized regulatory reporting templates are technically unnecessary in a fully granular model, standardized templates in a limited scope would provide a common, understandable view on the underlying data for both institutions and regulators and act as orientation points for communication and data visualization. Furthermore, it is of course possible to maintain the complete regulatory reporting template sets, especially regarding an inevitable transition period from a template-based to a contract-granular regulatory model or to maintain a long-standing time series.

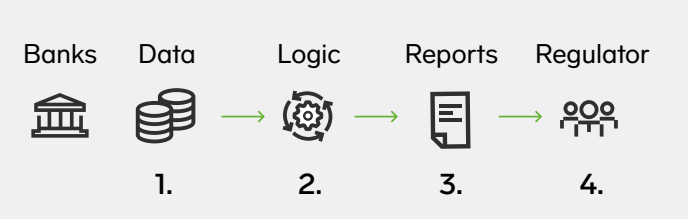


Figure 1: Typical regulatory process from a data's perspective

There are several initiatives to tackle the inefficiencies in the current regulatory processing logic. One of these projects is the "Digital Regulatory Reporting" (DRR) (FCA, 2017) project, initiated by the Financial Conduct Authority (FCA), to investigate the possibility of regulators to publish code as instructions and, by doing so, to overcome the additional efforts and costs to interpret and implement regulatory texts and definitions and to improve received data quality. This project has the potential to reduce the efforts of interpreting the regulatory logic but does not overcome the challenge of heterogeneous data sources and the challenge to map them.

End-to-end, integrated dataflow via API interface

The second element is an API, which would make the data formatted in the common data model and the regulatory logic accessible for outside parties, i.e., the regulators and supervisors. The API can potentially enable full access for the regulator to the complete data information on a granular level available at the supervised institutions. In this scenario, the complete, granular data can then be directly pulled (Bank of England, 2020) by regulatory authorities to their regulatory database via functional API requests from the bank's prepared regulatory data warehouses formatted in the standardized data model. Thus, the regulators can directly source data from the financial institutions via the API and use it for analytical purposes, possibly using big data and AI business intelligence tools.

The API's first function would be to provide a safe, fast, standardized portal or gateway to exchange regulatory data and processing logic. The API portal could be a central gateway that connects the financial institution's side with the regulator's side. A more likely scenario, especially in a multi-jurisdictional setup, would be multiple API portals, i.e., one portal per financial institution and one API portal for each regulatory authority. This enhances flexibility and resilience of the architecture and would allow the financial institutions to use the API portal also for non-regulatory communication with other external partners (business partners, API providers, third-party applications).

Another required function is the transmission and execution of regulatory requests and transformations. In this case, the regulator would formulate data requests not functionally but technically in code. This regulatory data request in code is published via a common code repository and then pushed to financial institutions, which can then execute it on their data warehouse and compile and process the requested data reporting package based on the common logic. The financial institution can then check and deliver the automatically generated data package to the regulator, or the regulator would automatically receive it. The bespoke reporting package could contain either template-based information, granular data, or both depending on the content of the regulatory data request. These requests would be flexibly useable to generate aggregated reports as well as gathering granular data.

Such an API-based architecture will increasingly enable the regulators to access the entire data information of the regulated entity more directly on a granular level while preserving the possibilities to execute transformation, aggregations, and template generation. The end-to-end data flow from uniformly standardized regulatory data warehouses and regulatory data requests via code for the whole financial markets will considerably shorten regulatory change cycles and provide a high degree of regulatory flexibility. Especially in the transformation period, the API also offers a way to build an IT architecture compliant with current legislation by allowing the financial institutions to control and sign off on data that is to be transferred.

Regulatory Big Data Platform

The third element is a big data-enabled SupTech solution, which could serve several purposes.

For one, it would be the tool regulators use to access the API and draw data to their own databases. In a regulatory reporting model with granular datasets sourced individually from each supervised entity, vast volumes of data would have to be transferred, collected, stored, and processed regularly. This solution needs to be natively built on a modern, big-data-enabled software stack to cope with data storage and data processing requirements. The solution must be architecturally suitable for modern technological frameworks for data formats/storage (e.g., Parquet, Amazon S3).

The second application is the processing of the functional units, i.e., the logic which controls the regulatory standard pull from the API, processes and transforms the imported data into the standard and ad-hoc formats required by the regulator. This also opens the possibility for the regulator to further process the received datasets with non-standardized algorithms and logic to increase the analytical depth. Data processing would have to be incorporated with high-performance, big data-enabled technologies for data processing (e.g., Apache Spark, Hadoop).

Furthermore, the solutions require an v that allows for the visualization of the data. This means the representation of standard templates and data views agreed upon for the market for reasons of transparency and communication. But it also means the possibility of creating individual dashboards and virtual templates that could be customized or created ad-hoc for information requirements the regulatory authority, department, or individual supervisory employee need for their specific supervision use case.

This toolset would need to be sophisticated enough to fulfill the analytical needs. Still, handling should be simple enough to minimize the need for consulting or IT resources for implementation/deployment. Again, the used technology would require sufficient performance for high data volumes for data visualization/analytics (e.g., Superset, Tableau, Power BI). Such a reporting system at regulators would be an ideal foundation for applying advanced BI tools or artificial intelligence algorithms on the data, but it is currently difficult due to the lack of common data standards and the low performance and the high cost of on-premise data handling capabilities.

The RegOps Model

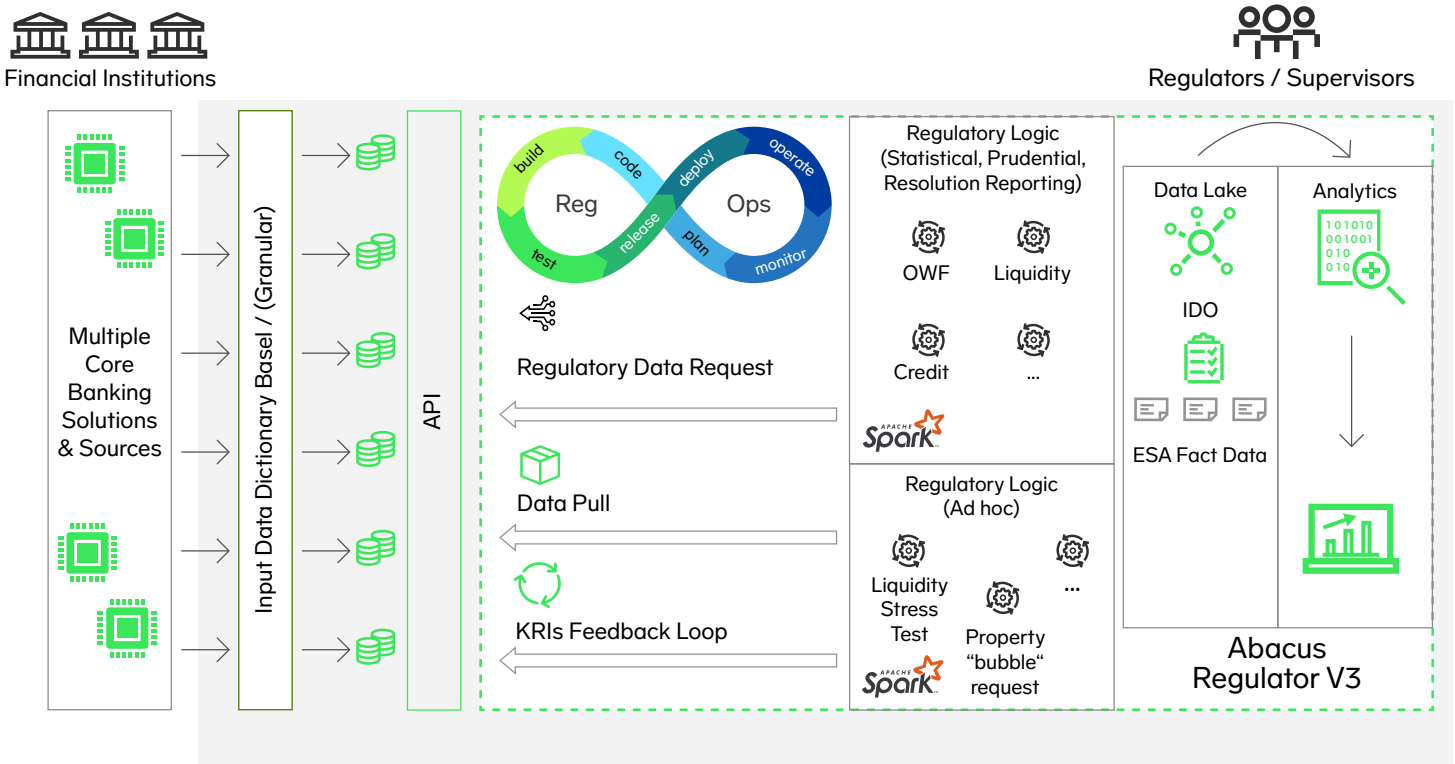
When we combine these elements, we can see a system where regulatory data is directly sourced from a highly detailed, complete, and fully granular standardized data model from every single institution. This data can then be accessed by a standardized logic to be imported, processed, and returned in standardized and ad-hoc formats for the regulator. The proposal would streamline the reporting stream as far as possible and largely solves the issues of system breaks in the current regulatory reporting flow. Furthermore, it would solve the issues of standardization for the data model and the data processing logic to ensure the highest possible quality and comparability.

An interesting side effect is that this model will virtually end the need for regulatory change on the side of financial institutions after a few iterations, as there is only a finite number of sensible information artefacts and data fields to be added to such a granular data model. The regulator can then flexibly build new regulatory templates without action required by the financial market. This is also confirmed by the authors' experience of operating granular data model-based regulatory reporting approaches.

The pull-based model could be gradually phased in to gain experience with granular data and give time to adopt regulation and legislation towards the new architecture while keeping the existing regulatory reporting infrastructure (push approach) in place to enable a smooth transition for all involved parties. If the model yields the envisaged benefits, the legacy push-based infrastructure could be migrated gradually to the new pull model, resulting in a scenario described in the diagram below.

Thus, we would see a natural shift of regulatory development from "waterfall supervision" to "agile supervision." This model would not only strongly increase the quality, timeliness, completeness, and transparency of regulatory reporting but also concurrently greatly reduce the cost and efforts for regulatory reporting for the affected financial institutions.

Standardized input data model	Regular	(X)
	Market	(X)
Pull-mechanism		X
Integrated data transfer		X
Granular data delivery		X
Big data-enabled analysis tool		X



RegOps Proof-of-Concept "G20 / BIS Techsprint 2020"
Regnology participated in the G20 / BIS Techsprint 2020 with the RegOps model and developed a fully functional prototype. The PoC was deployed by reusing existing software stacks and regulatory contents in a completely new manner. Regnology built a fully working demo version within a five-week time frame. The solution was shortlisted by the G20 / BIS Techsprint 2020 judges panel for the finalist round in September 2020 ((BIS 2020c).
The prototype was able to prove the feasibility of the model and to deliver the following feature set:

- A standardized input data model allows the whole regulatory value chain to speak a common, granular language (incl. Pseudo-LEI, Pseudo-UTI)
- Standardized, regulatory processing and allocation logic for Basel III CRE20 (Credit Risk standard approach), CRE22 (Comprehensive collateralization for credit risk standard approach), and financial reporting (European FinRep) available for all financial institutions in the system
- Automated processing, end-to-end data flow, and rapid deployment of regulatory updates by introducing regulation as code via a publicly hosted Git repository
- Two regulatory scenarios with two reporting dates each for two financial institutions incl. 1,6 million datasets and over 200 million pieces of information
- Automated creation of multi-jurisdictional reporting templates for CRE20 & CRE22 (Basel regulation Credit Risk Standard Approach & Comprehensive collateralization) according to Singaporean standards (MAS637) and European standards (COREP) and financial reporting according to European standards (FINREP) out of the same granular input delivery
- Drill-down into the contract-granular data underlying the created reporting templates via SQL queries
- Big data-enabled platform for regulators with granular analytics functionality via Business intelligence (BI) toolsets (Apache Superset was used for this proof-of-concept (PoC))
- Enabling all above features without any change requirement for the financial institutions after the initial data model mapping exercise

This leads to the following conclusions and benefits:

Improving efficiency, transparency, and stability of financial markets

- Leads to high-quality, granular regulatory data in a standardized data-model for deep data insights for forensic insights
- Standardized regulatory processing, allocation and validation logic strongly improve reporting quality and comparability
- Automatic, timely data routing and processing via an API- enabled the network to identify and forecast developments earlier and more accurately

High efficiency & robustness

- All components of RegOps are generally available, and initial deployment can begin within a short timeframe
- Regulatory code changes can be deployed rapidly by the regulators
- An indication of high-cost efficiency for financial institutions due to an expected significant reduction in regulatory change costs
- High degree of automation indicates a potentially highly operationally robust system

Open for extension & innovative tech

- Conceptually high flexibility and adaptability to varying deployment scenarios in different jurisdictions
- Conceptual compatibility towards upcoming technologies such as cloud computing, AI, DLT as well as legacy systems

Summary & Outlook

The RegOps approach is a novel approach to tackle the two most urgent issues in regulatory reporting today: the lack of quality of regulatory reporting data and the concurrently high, unsustainable efforts and costs of financial institutions to produce this data. These issues are rooted in a lack of granular data delivery, a standardized, common data model and processing logic, and end-to-end integrated data flow. These
are also the main inhibitors for the application of modern technology to the current regulatory reporting regimes and were confirmed by an FSB study when mapping to the expected benefits of regulators.

This paper identified core prerequisite features that a new regulatory reporting system needs to overcome the current issues. These are:

- Integrated data transfer
- Granular data delivery
- Pull-mechanism
- Standardized data model
- Big data-enabled analysis tool

The paper has demonstrated that many of these problems and features were already partially addressed via various approaches and initiatives by regulators, financial institutions, and solution providers worldwide and have shown that they are able to deliver positive results not only in theory but also in practice.

Furthermore, the authors have shown how an underlying IT architecture of such a system could be constructed adhering to the principles of modern data infrastructure architectures (Bornstein, Casado & Li, 2020).

The system would need to deploy standardized databases, which are the source systems for the regulatory system. These standardized source database systems require a common data model to incorporate a data dictionary, the data model itself, and a metadata model. This data model would need to apply to all kinds of supervised institutions and business models. Minimally, this model would need to be suitable for the regulatory requirements of the respective jurisdictions but should be designed, if possible, with supranational and international applicability in mind. The common root for such an integrated data model could be either a data model completely developed by the regulatory authorities, an extension of existing industry data models, or a hybrid of both. The authors argue that a hybrid approach that incorporates specific industry standards will most likely be the best approach for many jurisdictions where granular, regulatory data standards do not exist or where a focus lies on international and supranational applicability.

The second part of the IT architecture would be an API or multi-API setup able to route data from the granular, standardized regulatory source databases of the individual financial institutions to the database of the SupTech solution, which is deployed by the supervisory authorities. This API or multi-API setup will act as safe communication points and offer secure channels between the parties. Furthermore, API systems should deploy standardized, common functional logic for transformations and allocations of data accessible in code alongside the legal prose via common code repository applications. The API infrastructure could distribute calculations, transformations, or consistency/ quality checks decentrally, which are too computing intense for a supervisor's system. These operations would be stored in so-called functional units, which should be stable and publicly transparent for regular reporting requirements or non-public and flexible for individual ad-hoc requests. The API could also act as a buffer layer between the regulator and the financial institutions and could mitigate certain legal aspects concerning direct regulatory access to the databases of financial institutions.

The third component of the model is a big data-enabled SupTech platform deployed by the regulator to collect, store, and analyze the granular data reported by the financial institutions. This platform would need to be developed natively to handle big data volumes and respective tasks. It would also need to give the supervisors a toolset to analyze the data in a performant manner. The paper provides insight into which types of technology frameworks could be useful in this context. Furthermore, the SupTech solution would need to give the regulators the possibility to define standard or ad-hoc calculations or data requests. These so-called functional units can be processed by the API on the financial institutions' databases.

The authors argue that the first iteration of the proposed model is feasible with today's technology, available data standards, and current governance setups. For financial jurisdictions with common data standards, implementation could start instantly for a relatively low cost. It is realistic for other jurisdictions to envision an implementation effort of about 1-2 years for initial results if a common data model based on existing standards can be quickly agreed upon. Possible transformation scenarios give regulators and financial institutions ample time to learn about moving to completely granular data delivery by first applying the new system to experimental ad-hoc reporting (RegOps V1.0). Then they can begin to onboard the first legally binding reporting frameworks (RegOps V1.X) until eventually shifting all reporting requirements to the new system (RegOps V2.0). A specific timeframe for transforming the system is also needed to build up IT capabilities and train staff in financial institutions and regulatory authorities to handle and analyze granular regulatory reporting data.

Such a regulatory reporting system would mean a complete digitalization of the regulatory value-chain and solve many of today's problems. The shift from a regulation-driven to data-driven regulatory reporting is also a perfect base layer for the application of emerging technologies like blockchain (Münch and Bellon, 2020) (Regulation execution, data collection, and transmission), artificial intelligence (data validation, processing, and analysis), cloud computing (storing, processing) or quantum computing (calculations). Furthermore, a shift to a data-driven regulatory model would fit and facilitate the digitalization push, which can be observed for many data sources in issuance, trading, credits and loans, payments, KYC/AML, accounting, and other finance-related activities.

The authors urge regulatory authorities and supervisors to test new approaches to regulatory reporting and recommend conducting trials and proof-of-concept studies to validate approaches such as RegOps further. The studies should focus on finding a balance between what is technically possible and what is needed from a functional and governmental point of view for regulatory reporting. These trials should include all relevant stakeholders of the financial markets and not only regulatory authorities. It will also be highly beneficial if these trials are coordinated internationally, and information on the results are shared within the supervisory and financial market community. In a first iteration, the concept was shortlisted by the G20 / BIS Techsprint 2020 as a finalist solution for the future of regulatory reporting use case. It was subsequently implemented in a successful PoC project (BIS 2020c). The intention is to conduct further PoCs to learn about the implications of operating under the RegOps approach.

Literature

Bank for International Settlements, (2011). Basel III: a global regulatory framework for more resilient banks and banking systems. https://www.bis.org/publ/bcbs189.htm [Accessed 9th November 2020]

Bank for International Settlements and International Organization of Securities Commissions, (2017). Harmonisation of the Unique Product Identifier. https://www.bis.org/cpmi/publ/d169.pdf [Accessed 10th August 2020]

Bank for International Settlements, (2019). Basel III Monitoring Report. https://www.bis.org/bcbs/publ/d477.pdf [Accessed 10th August 2020]

Bank for International Settlements, (2020). The Basel Framework. https://www.bis.org/basel_framework/index.htm [Accessed 12th August 2020]

Bank for International Settlements, (2020b). Saudi G20 Presidency and BIS Innovation Hub update on the progress made in the G20 TechSprint initiative. https://www.bis.org/press/p200810.htm [Accessed 23th September 2020]

Bank of England, (2020). Discussion paper - Transforming data collection from the UK financial sector. https://www.bankofengland.co.uk/paper/2020/transforming-data-collection-from-the-uk-financial-sector [Accessed 10th Februar 2020]

Bašić, I., (2017). Supervisory and statistical granular data modelling at the Croatian National Bank (No. 25). ECB Statistics Paper. https://www.ecb.europa.eu/pub/pdf/scpsps/ecb.sps25.en.pdf [Accessed 6th November 2020]

Bornstein, M., Casado, M. and Li, J., (2020). Emerging Architectures for Modern Data Infrastructure. https://a16z.com/2020/10/15/the-emerging-architectures-for-modern-data-infrastructure/ [Accessed 6th November 2020]

Cœuré, Benoît, (2020). Leveraging technology to support supervision: challenges and collaborative solutions. 19th August 2020, Peterson Institute for International Finance. https://www.bis.org/speeches/sp200819.pdf [Accessed 1st October 2020]

Chartis and BearingPoint, (2018). Counting and Cutting the Cost of Compliance- How to accurately assess the cost of Risk Data Aggregation and Regulatory Reporting. https://www.reg.tech/files/Chartis_COC_Position-Paper.pdf?download=0&item=7197 [Accessed 24th September 2020]

Di Castri, S., Grasser, M. and Kulenkampff, A., (2020). An API-based Prudential Reporting System for the Bangko Sentral ng Pilipinas (BSP): R 2A Project Retrospective and Lessons Learned. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3596276 [Accessed 24th September 2020]

European Central Bank. (2019). The ESCB Integrated Reporting Framework (IReF): an overview. https://www.ecb.europa.eu/pub/pdf/other/ecb.escb_integrated_reporting_framework201902~83a269c171.en.pdf

European Central Bank. (2020). What is the BIRD? https://www.ecb.europa.eu/stats/ecb_statistics/co-operation_and_standards/reporting/html/bird_dedicated.en.html [Accessed 29th September 2020]

European Commission, (2019). Commission staff working document - Fitness check of EU supervisory reporting requirements. https://ec.europa.eu/info/sites/info/files/business_economy_euro/banking_and_finance/documents/191107-fitness-check-supervisory-reporting-staff-working-paper_en.pdf [Accessed 24th September 2020]

Financial Conduct Authority (FCA). (2017). Digital regulatory reporting. https://www.fca.org.uk/innovation/regtech/digital-regulatory-reporting#:~:text=Digital%20Regulatory%20Reporting%20(DRR)%20is,reporting%20more%20efficient%20and%20effective.&text=Regulatory%20reporting%20is%20vital%20to,and%20achieving%20our%20operational%20objectives [Accessed 29th September 2020]

Financial Stability Board, (2020). The Use of Supervisory and Regulatory Technology by Authorities and Regulator Institutions. Market developments and financial stability implications. https://www.fsb.org/wp-content/uploads/P091020.pdf [Accessed 6th November 2020]

Haldane, A., G., (2014). Managing global finance as a system. Speech at Maxwell Fry Annual Global Finance Lecture, Birmingham University. https://www.bankofengland.co.uk/-/media/boe/files/speech/2014/managing-global-finance-as-a-systemdf?la=en&hash=93BF6D650AAE5D055618D2D2DBC5870DC0580FA7 [Accessed 22th July 2020]

Kienecker, K., Sedlacek, G. and Turner, J., (2018). Managing the processing chain from banks' source data to statistical and regulatory reports in Austria. OeNB Statistiken, 3. https://www.oenb.at/dam/jcr:d9cdbe0a-a6d4-409a-8ac5-670cad2619b0/05_Kienecker_Statistiken_3_18.pdf [Accessed November 6th 2020]

McCrum, D., (2020). Wirecard and me: Dan McCrum on exposing a criminal enterprise. https://www.ft.com/content/745e34a1-0ca7-432c-b062-950c20e41f03 [Accessed 29th September 2020]

Münch, D. and Bellon, N., (2020). DLT-Based Regulatory Reporting – A game changer for the regulatory regime? SUERF Policy Issue, No 123. https://www.suerf.org/policynotes/9393/dlt-based-regulatory-reporting-a-game-changer-for-the-regulatory-regime/ [Accessed 22nd July 2020]

OpenFIGI, (2020). About FIGI. https://www.openfigi.com/about/figi [Accessed 22th July 2020]

Sanderson, R. and Crow, D., (2019). Jail terms for 13 bankers over Monte Paschi scandal. https://www.ft.com/content/54ace10a-023e-11ea-b7bc-f3fa4e77dd47 [Accessed 29th September 2020]

Van Steenis, H., (2019). Future of Finance Review on the Outlook for the UK Financial System: What It Means for the Bank of England. Future of Finance Report, Bank of England, https://www.bankofengland.co.uk/report/2019/future-of-finance [Accessed 15th July 2020]

Regnology
Speicherstrasse 1
60327 Frankfurt
Germany
regnology.net

Marketing/Sales Contact:
info@regnology.net
+49 69 567 007 910

About Regnology

Regnology is a leading international provider of innovative regulatory, risk, and supervisory technology solutions (RegTech, RiskTech, and SupTech), of AEOI and tax reporting products, as well as of services along the Regulatory Value Chain for financial services. Regnology has been a partner for banks and regulators for 25 years. Until end of 2020, the company was part of BearingPoint group and operated under the name Regnology. Since the sale of the RegTech business to private equity firm Nordic Capital, the company is independent. In June 2021, the company joined forces with Vizor Software and recently changed the name to Regnology. In total, Regnology serves more than 7,000 financial services firms with reporting solutions. At the same time, the company enables more than 50 regulators and tax authorities on five continents to collect data from 34,000 firms in 60 countries. The combined company has a total workforce of over 770 employees at 17 office locations in 12 countries.

More information:
www.regnology.net
www.vizorsoftware.com