

Regulations and Guidelines Monitor - Automatic Reminder for Regulations and Guidelines Updates

Dawei Jiao, Beijing KeyTech Statistical Consulting Co., Ltd;

ABSTRACT

As a practitioner in the clinical trial industry, compliance and enforcement of regulations and guidelines are crucial, and have become essential to our daily lives. Compliance with previous CDSIC standards is certainly important, but sometimes we are caught off guard when the standards are updated. In order to ensure sensitivity and real-time compliance with regulations, a software system established at KeyTeh has emerged.

Taking SDTMIG in the CDISC standard as an example, under normal circumstances, SDTMIG is updated irregularly. By monitoring web pages and combining them with the storage of the first Python backend database, information on web pages is obtained at fixed intervals and compared with the first data. If there is no difference, a reminder is given that there is no update or no reminder is given. If there is an update, relevant personnel are notified through the email system, and the update file is downloaded to the corresponding local path through the download system.

This approach saves most of manual monitoring work by 80-90%, and manual work free can be expected as the growing and recyclable metadata library being referred in future.

INTRODUCTION

Based on the fact that our industry has many regulatory documents. The previous CDISC standard regulations and guidelines have been well followed in KeyTech. While in the past, we have sometimes fallen behind in updating CDISC standards, regulations, and guidelines.

What's more, the AI technology is rapidly growing and with more widespread application in the industry, it came to us that it was necessary to build an intelligent tool -- Regulations and Guidelines Monitor (RGM), which aims to monitor CDISC standards and guidelines, provide reminders when they are updated, and automatically update them to a specific local path, and help us reduce manual monitoring work and downloading work, free us from these complicated and repetitive tasks to engage in more creative work.

Since RGM piloted in some standards update and reduced manual updating work significantly, this paper introduces how RGM is built, how you may use, furthermore, to discuss current limitations, challenges, and future expectations.

HOW RGM IS BUILT AND WHAT CAN RGM DO

As mentioned above, the RGM was developed with a back-end using the Python application and a front-end with cmd. The back-end uses complex algorithms to parse the metadata stored on web pages, and obtains the corresponding web page crawling and monitoring through the metadata. The front-end mainly uses the email system for reminders and the download system for downloading.

The Figure 1 demonstrates the overall process of RGM, and followings introduce key features:

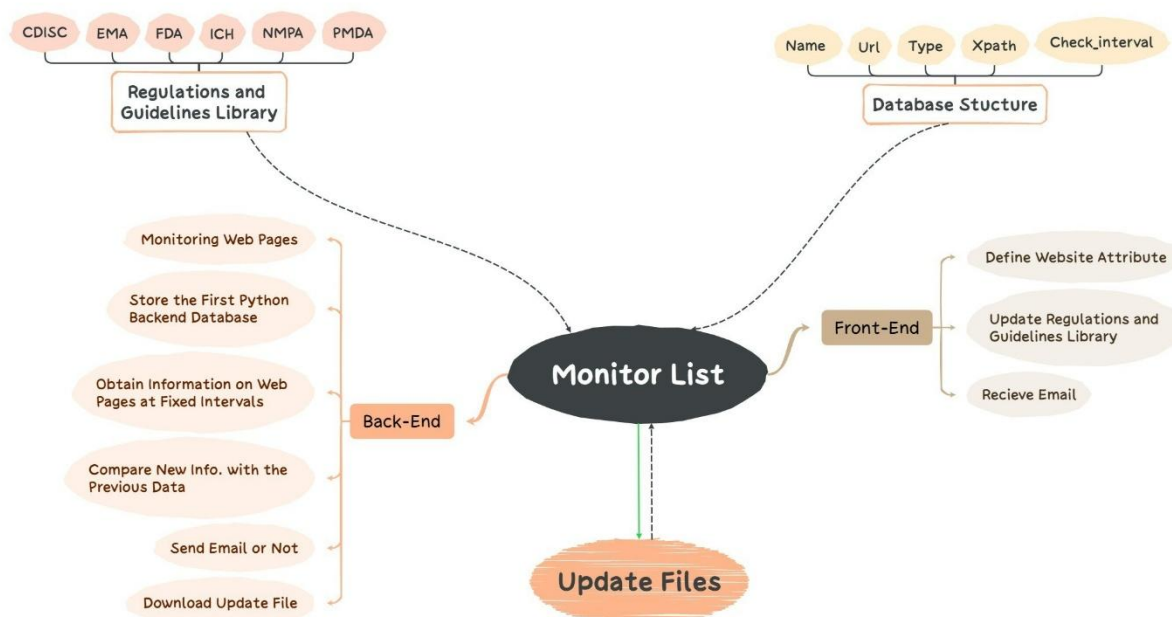


Figure 1. RGM Overall Process

1. COLLECT REGULATIONS AND GUIDELINES & INITIATE LIBRARY

The regulations and guidelines library is made up by at least 6 categories. Each category contains numerous regulatory URLs and documents, which form the foundation of RGM. The key concept here is to de-structure Website information as minimum components, which are easy for subsequent reference and assemble.

The regulations and guidelines library is initialized with the CDISC, EMA, FDA, ICH, NMPA, PMDA and other standards, take the SDTMIG in the CDISC standard as an example (Figure 2), the web page is de-structured as name, URL, and xpath three layers, each layer consists of information parsed directly web page, details stated below:

- Name:
 - Information directly from web page, like name or title or description of web page.
- Url:
 - Information directly from web page, defines as the URL of the website.
- Type:
 - The type of regulations and guidelines, like html, pdf, etc.
- Xpath:
 - The xpath of web page, information captured based on the html.
 - Some plug-ins or applications can help us to get this information.
- Check_interval:
 - The fixed time interval for RGM to monitor web page.
 - Unit: second.

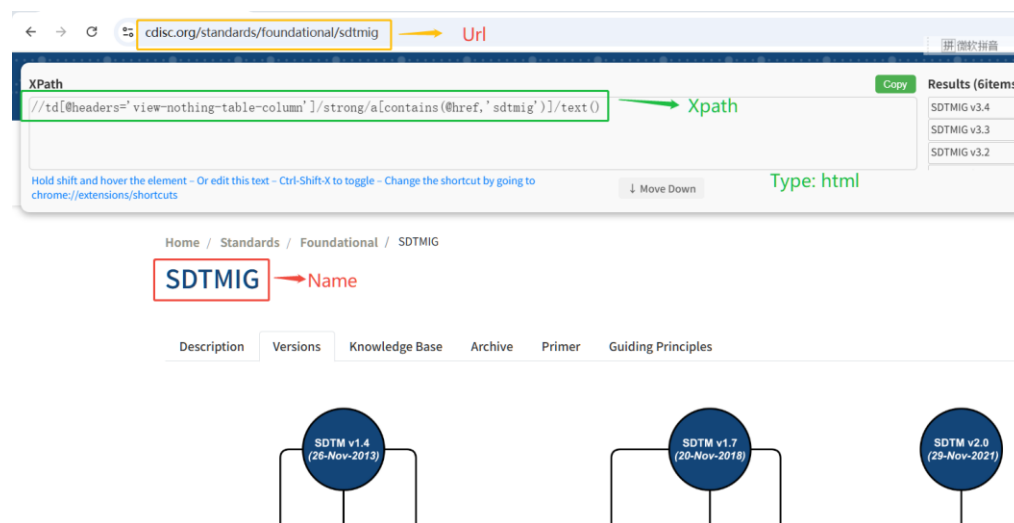


Figure 2. De-structured Web Page and Attributes

2. HOW RGM MONITOR WEBPAGES

So, what is a new web page comes to the RGM? First, RGM applies algorithm to de-structure and digitalize Web page contexts as metadata. Furthermore, RGM compares database structure with the reference metadata library, which one or more reference chose to refer. For cases that are match, nothing is going to change; for cases that are not match, RGM add the case into the library. Below are detailed and optimized steps regarding how to perform monitoring from input to output.

- 1) When the first webpage information is entered, Python's hashlib package will be used to obtain the hash value of the webpage and store the relevant webpage information in the database.
 - 2) After each time interval, Python's requests package will be used to obtain webpage information and new hash values.
 - 3) Compare the hash values of the new and old versions to determine if there have been any updates
- To achieve webpage monitoring through the above steps and methods.

3. HOW RGM SEND EMAIL

As above introduction, RGM digitalizes web pages as metadata, and also provide reminders when they are updated. which means once web page is updated, an email can be sent to us without any other operation.

Figure 3 demonstrates an email created for SDTMIG.

- Proper structure, consisting of email header, from who, and update content of website

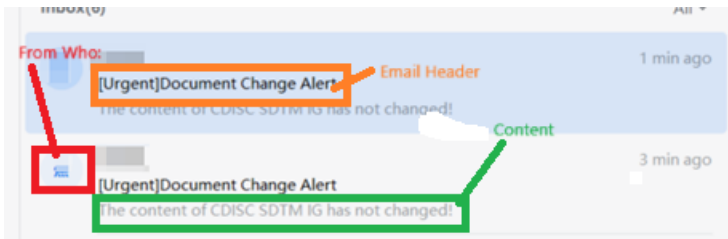


Figure 3. Email Sent by RGM

HOW YOU USE RGM

4. WHAT YOU NEED TO DO

Basically, you shall input some specific database attributes, upload web page information to create metadata and ensure metadata accuracy.

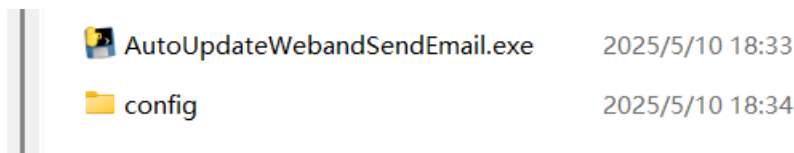
- Database attributes shall be input manually for now because they are not obtained either from the metadata library. Relevant library attributes include FDA or EMA standards, ICH and NMPA standards, etc. For example, URL and Xpath are generally different between FDA and EMA standards. Clarifying it helps conduct metadata matching with corresponding regulations and guidelines library and fetch appropriate attributes.
- You shall upload monitor list with json format to create digitalized metadata and get initialized attributes.
- You should ensure metadata are appropriate and with correct attributes, especially for URL and Xpath which are mapped through applications or plug-ins.

5. HOW RGM FACILITATES PROCESS AND REDUCES MANUAL WORK

So far, the software can only search for CDISC SDTM IG websites and send update reminders. Below is a simple demonstration using software, friendly user interface, and email sending. As mentioned above, metadata is the foundation of RGM to send email to us, but the purpose of RGM is not to transfer our work from manual monitoring to creating metadata. What RGM does want to achieve is making metadata automated and accurate, reducing manual work, facilitating metadata generation, matching, and finally monitoring and downloading update files. Apart from applying algorithm to improve metadata and match URLs, RGM does develop a user-friendly front-end interface to provide convenient operations and important hints.

Using Software

RGM is now a software and with a config folder. The config folder is the input of regulations and guidelines library. And once you update your input then the information can be stored into the database. And next you just need to let the Application run.



Display 1. Software Screenshots

Friendly User Interface

- First open the command prompt (cmd), then drag and drop the exe file into the cmd windows to run it. This way, you can view the defense alerts!
- Then you can first see the first option you can choose to input:

Please enter the monitoring interval in minutes:

This option is for you to enter the monitoring interval in minutes, if you do not input it, then RGM is going to use the default check_interval within the metadata.

- After that you can see the second option you can choose to input:

Please enter the recipient's email address:

If you do not input it, then RGM is going to use the default email address within the RGM, which is now my email address.

You may see the details in Display 2.

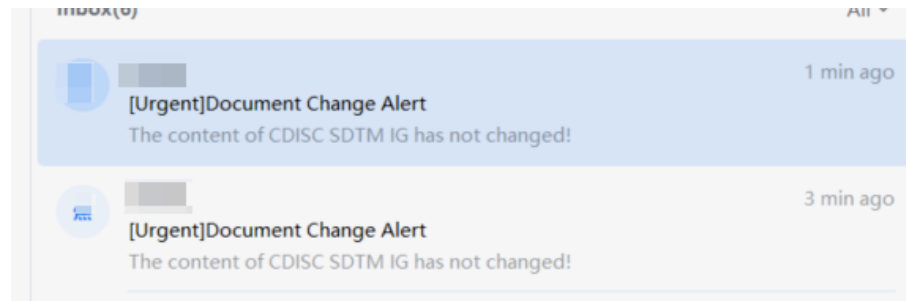
```
First open the command prompt (cmd), then drag and drop the exe file into the cmd window to run it. This way, you can view the defense alerts!  
Please enter the monitoring interval in minutes: 0.1  
Please enter the recipient's email address:
```

```
First open the command prompt (cmd), then drag and drop the exe file into the cmd window to run it.  
This way, you can view the defense alerts!  
Please enter the monitoring interval in minutes: 0.1  
Please enter the recipient's email address: Input timed out. Using the default email address.da.wei.jiao@ktstat.com
```

Display 2. Friendly User Interface Screenshots

Email Sending to Notification

As mentioned above, you can choose to notify you when there is update or not. The notification is sent to your email without any other operations. You may see the details in Display 3.



Display 3. Email Sending Screenshots

FUTURE DIRECTIONS AND CHALLENGES

There are some plans and challenges as follows:

- Monitor all standards: Monitor all websites related to our industry to achieve the initial goal.
- Automatic monitoring: Currently, manual clicks are needed to use the software, and automatic monitoring is not yet available.
- Add download part: So far, only email notifications are in place. The next step is to automatically download updates to a specified location.
- Enhanced multi - level warning notification: AI has suggested a multi - level early warning function, which we haven't considered in detail yet. It will be a future consideration.

- Enhanced AI application: With the development and application of AI technology, RGM also plans to explore and utilize more AI models to optimize back-end algorithm and enhance the capability, like large language model. Taking advantages of AI to allow us focus on something more valuable.

CONCLUSION

This paper provides the end-to-end Standards automation update implemented in KeyTech, introduces concept and product, discusses advantages and limitations of RGM.

Current RGM is founded with basic functions and requires some necessary user interaction. With further exploration, development of AI methods and utilizing novel technology, RGM will continuously iterate and optimize to become more user-friendly, powerful, automated, and intelligent.

REFERENCES

Kimi [AI chatbot]. Moonshot AI; 2024. Accessed May 10, 2025. from <https://kimi.moonshot.cn>.

DeepSeek [AI language model]. DeepSeek Inc.; 2024. Accessed May 10, 2025. from <https://www.deepseek.com>.

"PyPI · The Python Package Index". PyPI. Accessed May 10, 2025. from <https://pypi.org/>.

ACKNOWLEDGMENTS

We want to thank Nan Li, Jinting Luo, Yingwei Si and Zhiwei Jiang for the support on this work. We also want to thank Xuan Yang and Chengeng Tian for their reviews, comments and support.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Dawei Jiao
Beijing KeyTech Statistical Consulting Co., Ltd
da.wei.jiao@ktstat.com