GitHub
Source Control for Laboratory Digital Transformation

1. Introduction

As a scientific and technology leader, Thermo Fisher develops and implements industry-leading LIMS (Laboratory Information Management System) applications. Industry and market-leading organisations partner with Thermo Fisher to modernise and digitalize their laboratory operations with LIMS. This paper discusses the added benefit of introducing a standardised approach to software source control using GitHub. As well as supporting customer validation efforts, Thermo Fisher can also provide full traceability of codebase, configurations and releases. This builds trust on the foundations of responsible implementation practices, software and project delivery governance. Thermo Fisher also benefits by building a catalogue of artefacts that can benefit the core LIMS products and provide insights into repeated work and effort by global delivery teams. As Thermo Fisher is engaged to deliver a LIMS to a customer on-premise system, customer hybrid (cloud and on-premise) or hosted system it also discusses the considerations taken for infrastructure as well as software and how artefacts from both are considered for storage in source control.

1. Research Methodology

This section highlights the current use of GitHub by Thermo Fisher development teams. This provides evidence of the current state of pull requests and variations of programming languages that are used by other teams supporting the proposal of the DSS (Digital Science Solutions) team to adopt the use of Github.

1.1 Research Data

Data was extracted directly from Thermo Fishers' Github corporate account 'insights' dashboards. The dataset includes the annual trends for pull requests across all repositories and the total count of primary language used in repositories.

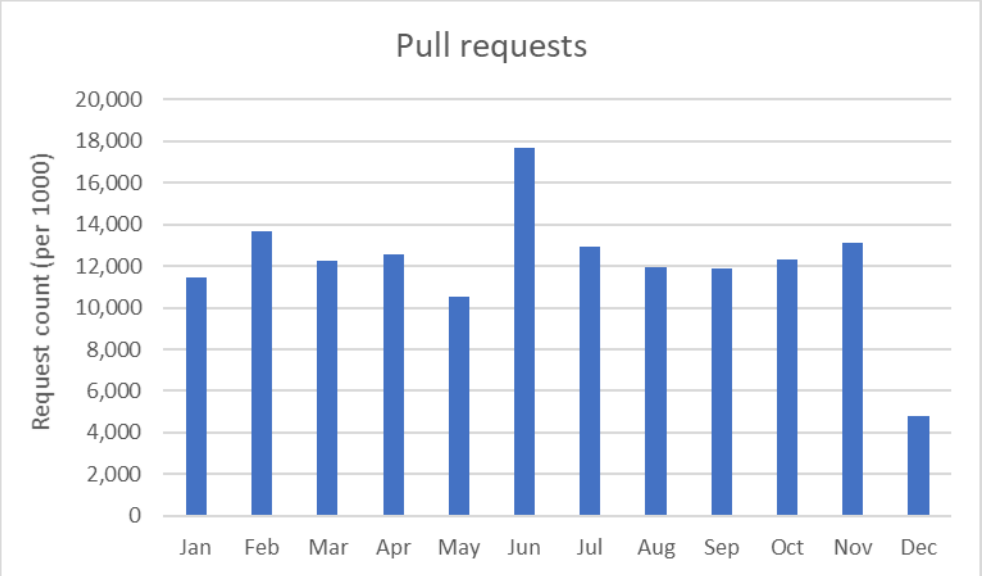

Fig1. Illustrates the existing utilisation of GitHub by Thermo Fisher development teams and the pull requests for existing repositories over the course of 2022.

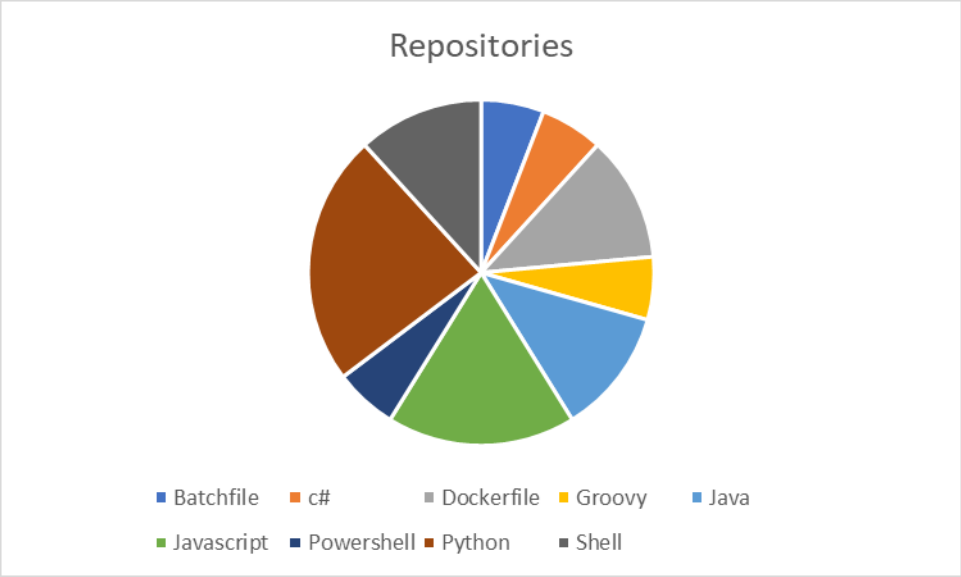

Fig2. Illustrates the mix of languages used in existing repositories.

2. Drivers for adopting Github

Organisations put the digital transformation strategy at the forefront of their business when building out their technology platforms. Scientific-based organisations regularly look to implement a LIMS to partner with their traditional and established enterprise systems.

At the heart of these strategies changes. Change to process and change in mindset from both the customers' perspective and Thermo Fishers' are key in understanding how the two parties partner with each other to work towards the customer objectives. Baker (2015) talks about if the change is inevitable then it will be transformative but risks introducing disruption and resistance. By adhering to a standard tool like GitHub, Thermo Fisher can engage its internal teams and foster acceptance of a new standard which builds trust with all stakeholders. Our research shows that Thermo Fisher has a well-established use of Github, therefore, has the resource and skills to support the DSS team to adopt its use. Blischak, Davenport & Wilson (2016) present an introductory discussion on the use of a VCS (Version Control System) where they discuss the benefits to individuals from both the scientific and programming fields. As Thermo Fisher's customers are predominantly science-based organisations and its team are technology professionals or from a scientific background we can see the many benefits of utilising source control.

There is also a clear distinction to be made that LIMS implementation projects are based on a COTS (Commercial Off-The-Shelf) package and are not considered development or control of the core product itself but rather the extensible configuration. Alongside the core application, infrastructure and architectural requirements there is a heightened interest in how the configurations of the software and hardware are recorded and stored to ensure quality and what has been released.

Fox, Lantner & Marcom (1997) discuss the evolution of applications and how they separate themselves into layers beyond the infrastructure and core solution. This is another key area of consideration when defining the source control strategy, as we are not as concerned about the compiler or root level of the application but rather the

artefacts that define its use. This can cover user accounts, reports, workflows and calculations that are configurable. Any source control system should also consider these artefacts in addition to traditional concepts, such as .net code and compiled DLLs. The more complex the strategy, the more risk there is to the management and delivery of the project. Hassan & Holt (2003) discuss the chaos of software development and how complexity is also a contributing factor to the lifecycle of a development project. If Thermo Fisher introduces too much complexity to their approaches, then they could surmise that the primary goal of a standard global approach may not be met. This could also introduce significant challenges to in-progress projects, where time and resource is spent on the management of poorly defined and managed repositories versus the delivery.

Historically, Thermo Fisher services teams have worked with very limited control or management around software repositories. Until the more recent introduction of Thermo Fishers' cloud-hosted LIMS, customers traditionally build infrastructure in their domain. Thermo Fisher consultants connect to these environments through controlled access and adhere to the customer's IT standards. In some cases, customers show no preference or governance about how they want their implementation to be managed. Thermo Fisher has developed many processes. This includes project management, development methodologies (Agile and Waterfall), documentation and QMS (Quality Management System) processes but no solid approach to source control. Thermo Fisher has also implemented standalone GitHub subscriptions on a project-by-project basis. Alternatively, customers can and will mandate the use of their source control systems. Thermo Fisher proposes to use their own corporate GitHub subscription as the default offering which in turn introduces consistency and standards that form part of our delivery, quality and responsibility.

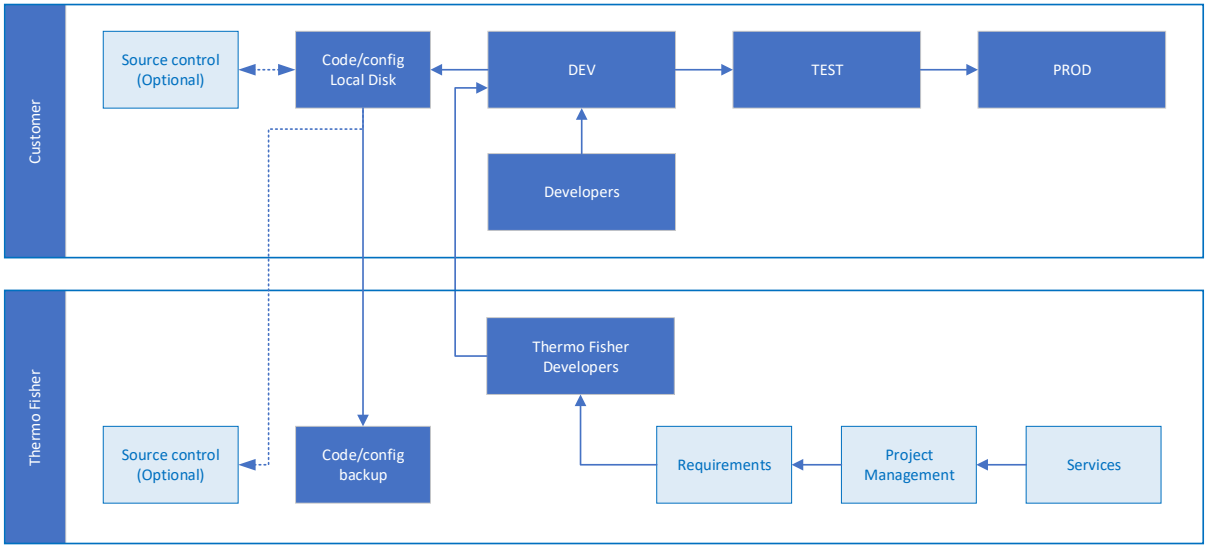

Fig3. Illustrates a traditional project deployment where configuration outputs often live locally or as a backup, but versioning is manual and dependent on the releases and methodology.

As discussed previously, Thermo Fisher provides a cloud-hosted LIMS. This managed solution has been a primary driver behind the decision to implement a VCS

strategy. As well as the traditional artefacts from a COTS development package there is also the need to capture artefacts from AWS (Amazon Web Services). Several artefacts are built or configured beyond the standard architecture which sits between software and infrastructure. Thermo Fishers' PaaS (Platform as a Service) solution gives the services teams the flexibility of automation in deployment scripts as well as custom management of AWS services. For example, in an on-premise solution, a customer may define views or indexes in the relational database and will own the responsibility for maintenance. However, in AWS an RDS (Relational Database Service) database may need configuration implemented by an infrastructure team. This, however, forms part of the overall project delivery, so we also need to consider infrastructure artefacts as part of the source control and release cycles.

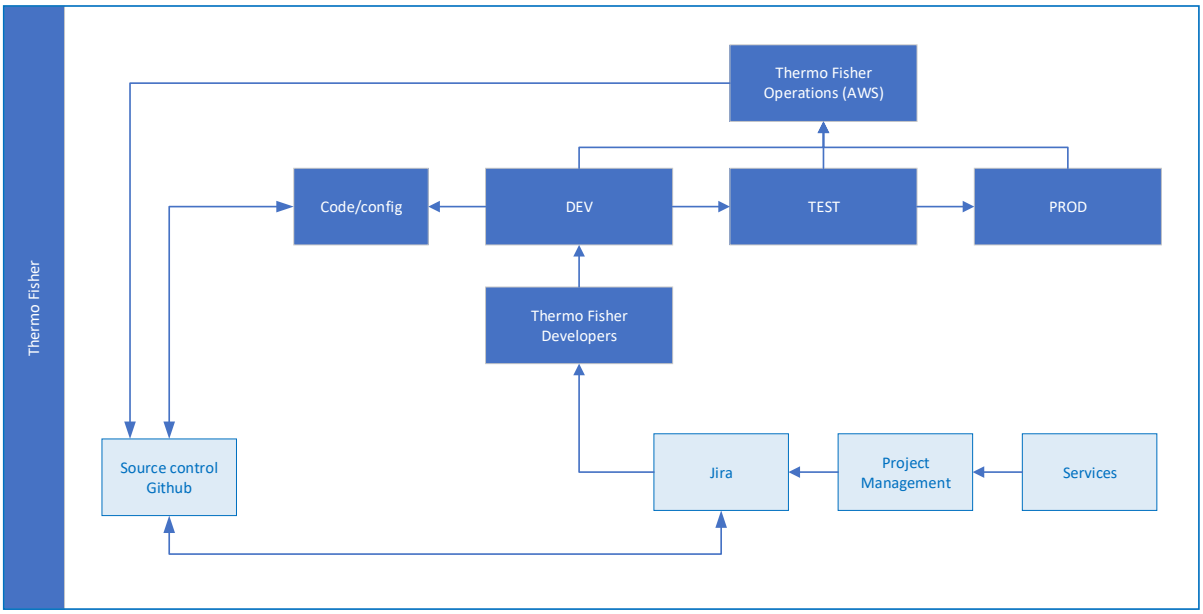

Fig4. Illustrates a hosted implementation where we see development activities and artefacts managed by GitHub. Additionally, we capture artefacts from technical operations staff recording infrastructure-level changes. Integrations between Github and Jira provide visibility against requirements and what has been built.

3. Advantages of a standardised strategy

As we consider the topic so far, it leads us to discuss the advantages of using VSC. As part of any implementation project, Thermo Fisher adheres to a range of quality assurance practices for the product, services, processes and people. As a customer continues their digital transformation, Thermo Fisher takes the responsibility of managing these practices to ensure quality and consistency. The configuration of the LIMS system to meet the customers' functional requirements is the stage the benefits of source control and how it is managed come to the fore. Capturing all artefacts through each release and build cycle, whether it be an Agile or Waterfall methodology that has been adhered to, builds a repository for a given project. We then have traceability between requirements, and specifications (functional and technical) and align with tools like Jira. This gives Thermo Fisher a valuable set of insights into the project delivery with internal governance and valuable reporting tools for both internal and external stakeholders. Furthermore, by leveraging tools

within GitHub itself, Thermo Fisher can benefit from automation and integration tools. The organisation in a future state can look to integrate with PSA (Professional Services Automation) and CRM (Customer Relationship Management) tools to give further visibility to the commercial aspects of the organisation.

The advantages of using Github, further highlight the impact on Thermo Fishers' corporate and social responsibility, both internally and externally. It positions the company for a shift to using an industry-standard platform and empowers the delivery teams to deliver high-quality products. This does lead to some assumptions that need to be made on the individual contribution. Whilst having a software platform that can record and capture changes to a software product, we also begin to understand the potential risk of an individual's personal or professional attitude and commitment to understanding the value. Guzman, Azócar & Li (2014) discuss emotion, time, and individual and team contributions as factors that can influence the commit comments used in Github. A programming language, a day of the week or project methodology may influence the individual or team to perhaps leave unnecessary or controversial commentary. While this is a risk, if Thermo Fisher encourages engagement and provides training the organisation can collectively understand the benefit of using Github and manage the change accordingly. It is, after all, the metadata of Github as well as the artefacts themselves that offer the real value to both the external customer, their digital transformation and the internal stakeholders. Another potential benefit is the possibility to encapsulate the standards implemented in Github to conform to ISO 9001. If we consider configuration management to cover both the development and delivery work as well as the collation and storage of artefacts from those activities, then one could consider adopting and adhering to a globally recognised framework and standard. Bamford & Deibler II (1995) summarise the definitions of CM (Configuration Management) as defined by the ISO and IEEE. GitHub and its contributors should be able to identify, control and track versions of software artefacts by using the mechanisms within this framework.

4. Security and best practice

A key aspect of implementing this strategy is the security of the system and conformance to Thermo Fisher's corporate guidelines and best practices. The DSS team consists of project managers, business analysts, developers, consultants, as well as contracted partners. It cannot be assumed that Github or any other supporting system would have open access to these individuals but rather a granular security model to allow or restrict access based on role and responsibility. Thermo Fisher adopts a tightly controlled security model for systems like Github which includes the use of MFA (Multi-Factor Authentication). Further security controls then dictate the more traditional administrator, read or read & write as examples to ensure that teams and members of GitHub are controlled in what they can do. Managed by Thermo Fishers' corporate information technology and security teams, this compliments the quality procedures already in place at Thermo Fisher. Another benefit is the control of sensitive data such as keys or passphrases that might be used in certain configurations. Utilising settings in Github we can also ensure these items are stored securely and do not remain as static file-based objects residing on an operating system.

5. Conclusion

There is change required for Thermo Fisher and with appropriate support, engagement and encouragement from its leadership, it positions itself to further enhance its reputation in a very competitive market.

By leveraging an already well-established VCS and defining a robust design and strategy, the DSS services organisation can extend existing tools and processes by maintaining valuable configuration artefacts. In doing so, they become responsible to their customers and will continue to build on their reputation as an industry leader.

References:

Baker, M. (2015) Digital transformation.

Blischak, J.D., Davenport, E.R. and Wilson, G. (2016) A quick introduction to version control with Git and GitHub. *PLoS computational biology*, *12*(1), p.e1004668.

Fox, G., Lantner, K. and Marcom, S. (1997) A software development process for COTS-based information system infrastructure. In *Proceedings Fifth International Symposium on Assessment of Software Tools and Technologies*: 133-142. IEEE.

Hassan, A.E. and Holt, R.C., 2003, September. The chaos of software development. In *Sixth International Workshop on Principles of Software Evolution, 2003. Proceedings.* (pp. 84-94). IEEE.

Guzman, E., Azócar, D. and Li, Y. (2014) Sentiment analysis of commit comments in GitHub: an empirical study. In *Proceedings of the 11th working conference on mining software repositories*: 352-355.

Bamford, R. and Deibler II, W.J. (1995) Configuration Management and ISO 9001. *changes*, *2*: 3.