ontotext

A great example of how knowledge graphs have helped an industry advance their work is clinical diagnostics. Clinical diagnostics labs help clinicians diagnose, treat, and manage patients by testing and analyzing human specimens.

As these patients are waiting to see if they have potentially life-altering issues – like cancer, for example – and results quality, integrity, and repeatability are crucially important to these labs.

Laboratory Software

To promote efficiency and accuracy, most clinical laboratories use software like a Laboratory Information System (LIS) and/or a Laboratory Information Management System (LIMS) to manage activities and data in the lab. With low margins in a highly regulated environment that requires personnel with specialized training, they have traditionally invested minimally in this software, prioritizing the hiring of specialists from a relatively shallow pool of talent.

In use for decades, LIS and LIMS software have typically been built on relational databases with rigid data models. And up until recently, the lab tests were relatively simple, point-in-time snapshots of a single quantitative result. Laboratory testing workflows were easily satisfied by this software.

Around 2015, Next-Generation Sequencing (NGS) became an accepted diagnostic tool with data capture that was more complex than a simple point-in-time snapshot. It required the ability to capture the entire story of a sample's journey through a lab. It is at this point that the rigidity of the relational database began to present shortcomings for clinical diagnostic laboratories.

Clinical diagnostic labs – a functional view

To provide context, let's explore a simplified view of what clinical diagnostics labs do. They receive a tube with a human specimen, typically blood or saliva. They accession it into their system, which is a formal way of saying they take custody of it. Then they do a quick quality control check to ensure there is enough specimen to process the sample.

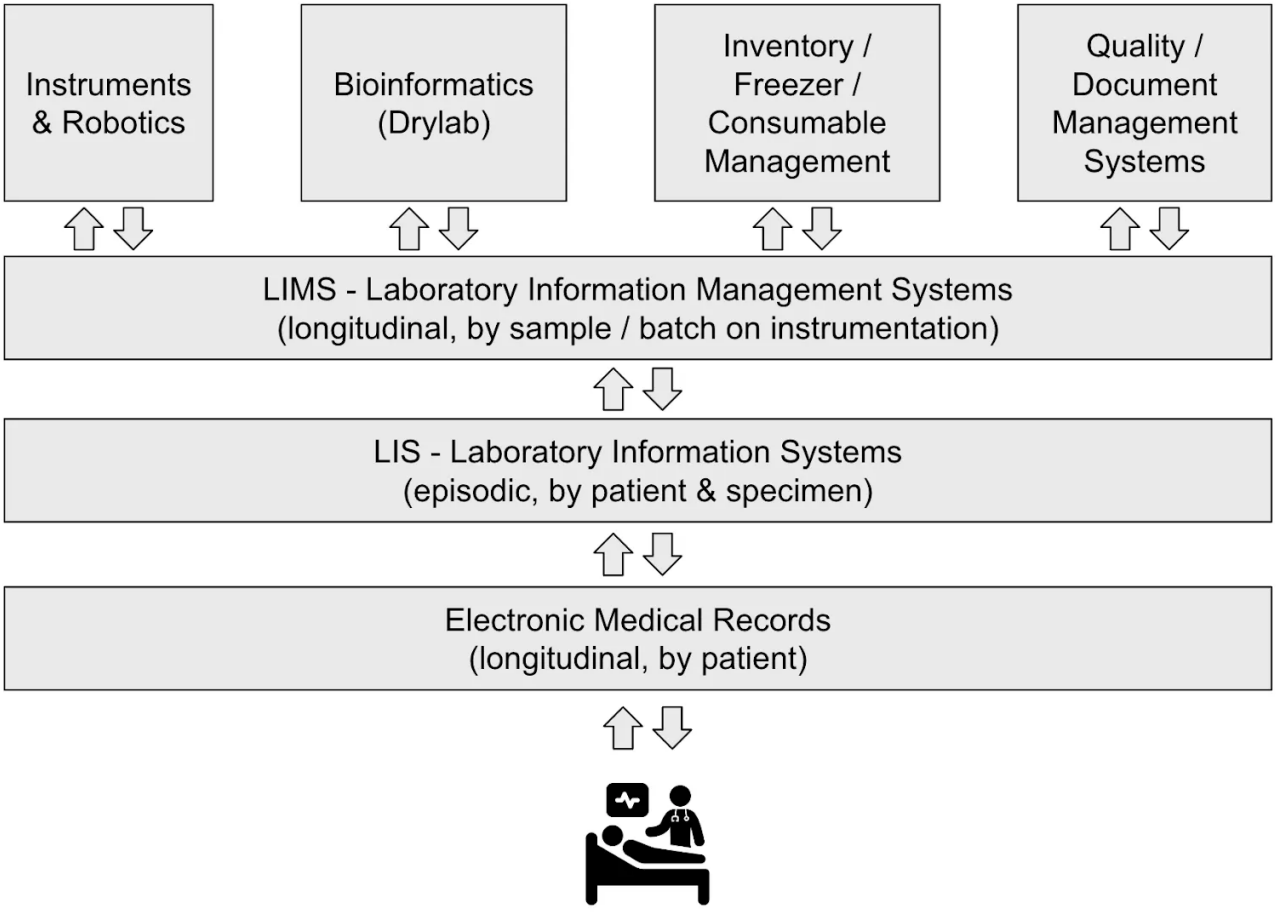
Once this is done, they set up a preparation process to go to an analysis machine. Then they

Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com

analyze it with the relevant tools and process the analysis into a result. Finally, they create a report, sign it (ideally electronically), and that's it. Sounds easy, right? Well, the reality is different.

Clinical diagnostic labs – a systems view

The following diagram shows a patient and a stack of 7 or sometimes even 12 software tools that together produce a lab result.



There are a lot of enterprise vendors who have developed point solution software to accomplish a particular task. For example, LIS are patient-centric, while LIMS are sample and

Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com

batch entity centric. So labs have to buy, integrate, and learn all of these systems to operate what is ultimately a low-margin business in a regulated space.

Looking specifically at NGS, before a novel therapy can be administered, it requires a new diagnosis. And each new diagnosis pathway requires changes to every system in the diagram. Each system has to be reconfigured, revalidated, and released through an interdisciplinary process. And, this is only what's needed for implementing it, not improving it over time. In our experience, that release process with systems based on relational data stores can take 3, 6, or even 12 months depending on how big the change is.

Unfortunately, labs with systems architected around a particular relational data model are stuck with this slow process which hinders their ability to rapidly deploy improvements. Additionally, all process elements like patients, samples, inventory, freezers, and people are tracked separately in the disparate systems, making it very difficult to build a complete picture of a sample's journey through the lab.

Knowledge Graphs and Life Sciences

The only way to naturally model Life Sciences data and preserve the value of the work that regulated labs put into their data is a knowledge graph. And, from our perspective, Ontotext offers a truly performant RDF knowledge graph. It provides multiple benefits to these labs, which produce data that will unlock the next frontier in human health. But that can happen only if they can start storing that data in a reusable way rather than distributing it across multiple normalized relational database structures.

In my opinion, what prevents these labs from accessing the benefits of a knowledge graph comes down to inertia. Every software engineer I talk to outside our company still defaults to relational databases. It's understandable. Where is the incentive for a 5, 10, 20, or even 30-year-old product, selling into a low-margin space, to fundamentally re-architect their application and adopt a knowledge graph? So, usually, a knowledge graph gets foisted upon these folks because of a FAIR-ification initiative or something similar.

Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com

ontotext

Embarking on a new path

Back in 2018/2019, we at Semaphore, as professional implementers of 3rd party LIMSs, got fed up with trying to abuse the older brittle lab platforms into a shape that could serve our clients' needs. Instead, we embarked on a new product development journey that we envisioned as the future of these labs and decided to chart a path to get them there.

In this future, lab data can be naturally modeled in the software in the same way the humans in the lab think about it. This would remove the interdisciplinary vocabulary barrier between software data science and laboratory science. It's a future where the incredibly rigorous nature of data capture is not only defined prospectively in advance but also fully captured retrospectively.

Labbit's conceptual core

When we started to develop our laboratory informatics platform Labbit, we came up with a high level of requirements, which are applicable across many industries:

- Capture the data
- Control the process
- Contextualize historical data & its source (provenance)
- Easy to use
- Integrated where needed
- Able to improve the process over time

After a lot of consideration of how to best achieve these requirements, we decided to adopt the FAIR data principles as baseline system requirements. We also agreed to use an RDF knowledge graph as our database and chose Ontotext GraphDB for the job. That obviated many integration data issues we would have experienced. At the same time, it future-proofed both our and our clients' investment for the coming large language model revolution.

To ensure that everyone could trust the data produced, we layered a provenance center W3-Prov throughout it. In terms of process control, we discovered the ISO standard business process model and notation (BPMN) had two advantages. One, it naturally mapped into W3-Prov, and, two, we could visually model almost any process or process control in it, and make

Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com

ontotext

that model semantically executable. Out of the numerous BPMN workflow engines out there we chose Camunda to control the process flows in our platform.

We also blended in some modern software development. We did deep user experience research about what these lab folks were expecting independent of the discipline we were bringing to bear. We used infrastructure as code for instantaneous deployments containers and microservice-style architectures to make the platform horizontally and vertically scalable. In addition, where the platform expects to be part of a modern development pipeline, it supports and enables these steps. All this makes the system incredibly flexible, efficient, and scalable.

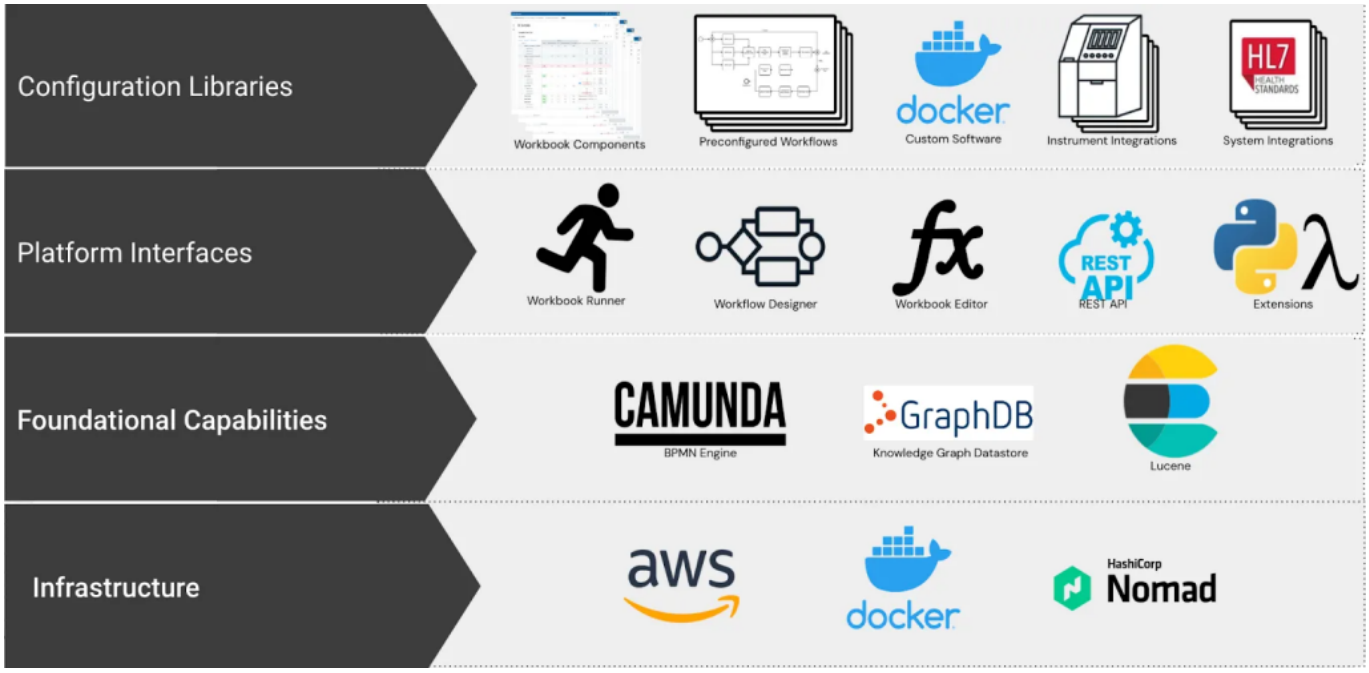
Our final step was to think about how to help our clients manage changes to a system, built to be as flexible as the knowledge graph, in a way that would speed things up. So, we created an entity that refers to a complete set of changes to the system that can be deployed to any virtual environment. Since that “change set” is an entity, it can be controlled by a workflow. Every entity in the system is stamped with which “change set” it went through at production before it became retrospective data. This enables a machine to reason about what real-world process it went through and to trust that process.

Labbit’s architectural core

The following high-level diagram shows what everything looks like in practice. It’s not architecturally accurate but shows how we stack the elements together.

Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com

ontotext



Our value proposition and impact

Initially, we built our platform to fill in a gap that we as third-party implementers felt. We saw that in most labs the software architecture was an agglomeration of tools, which created a lot of issues, especially when they had to implement big changes.

So, instead, we built a system that is agnostic of control and could work with any other system out there, pulling its capabilities into it. Because of Ontotext's great work, our platform, by design, can capture and quantify anything. When designing software that runs on top of a knowledge graph, you have to build it to be as flexible as the graph. This means that our clients can change anything they want in our system.

And, importantly, because of the work we did in prospective process control and chain history, we've created a system that not only lets you model your lines of business. It also allows you to model the meta workflows. So, the lab's higher-level workflows are all in the knowledge graph and it's easy to surface in real-time critical information that would be useful to an operator.

Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com

We often hear that the pace of innovation is directly related to the pace of iteration or experimentation. Now, with a laboratory system that lets you both define and control all of the changes to it while helping you make them smaller and faster, you can dramatically increase your iteration rate. As a result, the innovation rate we so desperately need to advance human health also increases.

So, now our clients can move from doing deployment once a quarter to once a week (or ideally even less than that) while staying compliant with all validation and regulatory requirements.

Challenges and the path forward

Every single clinical diagnostic lab is under-resourced and overworked. They can't compromise on quality, integrity, or repeatability, so something else has to give. Usually, it's seeking out tools like knowledge graphs.

On top of that, every lab runs on a tight budget. Each of the multiple tools in their software stacks can cost hundreds of thousands of dollars and sometimes requires 3 to 6 months to implement. It means these tools can't be easily replaced no matter how well they fulfill their purpose. Using 5+ unique scientific software tools also results in very siloed data and failed FAIR data initiatives. When people are already on the edge of burnout, they don't want to learn or invest their time in such initiatives.

So, the path forward is to recognize the value of these labs as, in my opinion, the data they produce is some of the highest-quality data in the world. The best way to recognize that value is by respecting their time and we can do that by reducing the jargon. These folks already have a Ph.D. in genetics, or molecular biology, or computational biology. Don't ask them to additionally learn the jargon associated with a whole other field just to collaborate.

Another step is to provide opinionated ways of doing things. When you put software out there, it should have a clear opinion about how to accomplish something but it shouldn't be so opinionated as to be constraining. And, finally, you need to slow down and not push big-bang initiatives. Whether it's a FAIR-ification effort in a 20,000-person company or a new software tool in a company of 50 people, you need to slow down, document, and implement.

Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com

ontotext

Main Takeaways

In the clinical diagnostics industry, where precision, efficiency, and repeatability are paramount, traditional laboratory informatics systems like LISs and LIMSs based on rigid relational databases have become inadequate for modern diagnostic tools, particularly NGS. Software built on knowledge graphs, like Semaphore's Labbit, offer a solution by capturing the complete journey of a sample, including its complex and deeply interconnected relationships with a variety of elements in the lab.

Do you want to learn more?

Book a Discovery Call

Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com