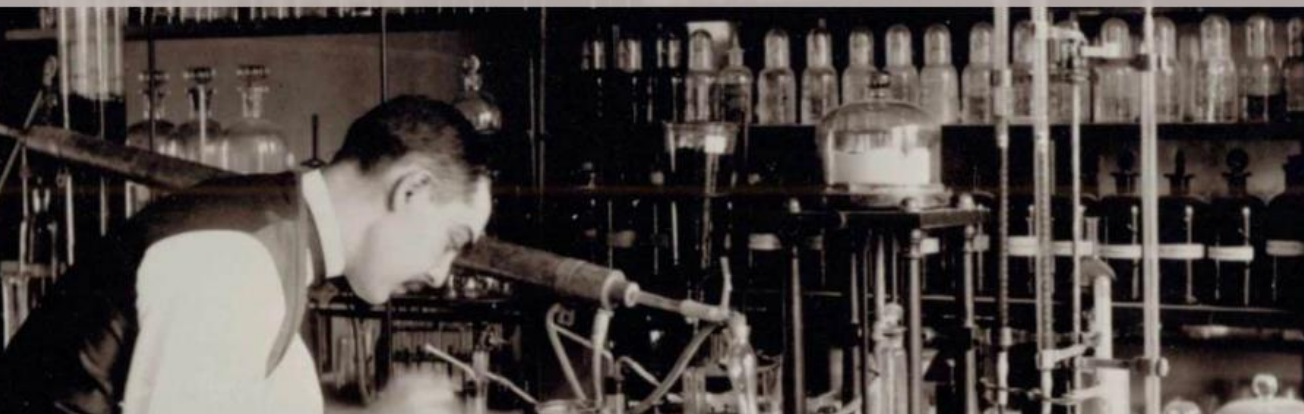


James A. Lumley, Research IT, Eli Lilly



# Deploying KNIME to the Enterprise Reshaping Data & Architecture for Healthcare





We were founded in 1876 by Colonel Eli Lilly, a man committed to creating high-quality medicines that met real needs in an era of unreliable elixirs peddled by questionable characters. His charge to the generations of employees who have followed was this: “Take what you find here and make it better and better.”

More than 140 years later, we remain committed to his vision through every aspect of our business and the people we serve starting with those who take our medicines, and extending to health care professionals, employees and the communities in which we live.



## OUR PROMISE

Lilly unites caring with discovery to make life better for people around the world.



## OUR MISSION

We make medicines that help people live longer, healthier, more active lives.



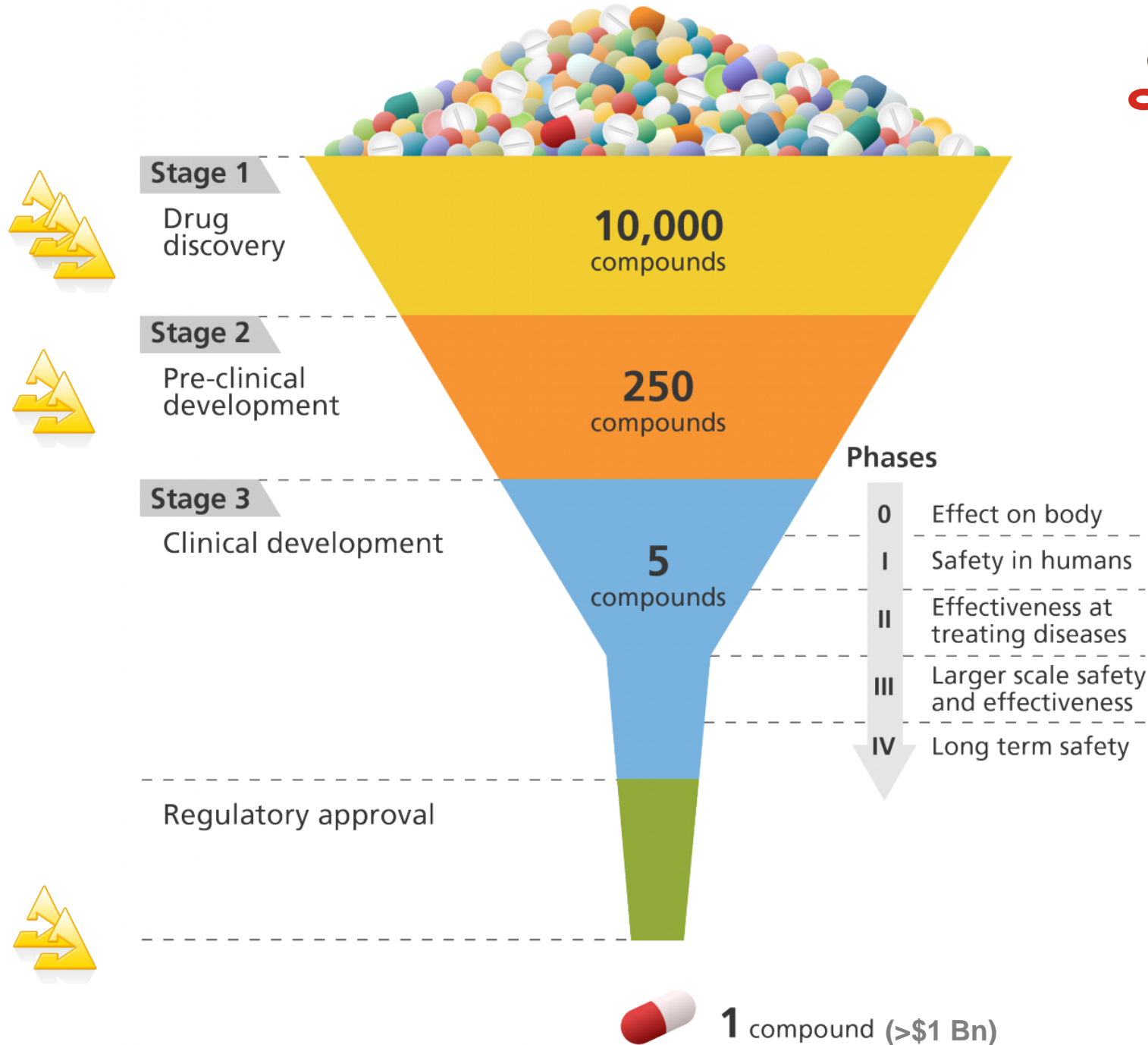
## OUR VISION

We will make a significant contribution to humanity by improving global health in the 21st century.



## OUR VALUES

Integrity, excellence and respect for people.



## Stable

Reliable,  
Reproducible,  
Transparent for  
development  
work

## Secure

Providing the  
right data at the  
right time to the  
right people

## Specific

The right  
analytics for my  
specific  
business  
problem

## Scalable

Can it handle  
the volume and  
variety of our  
data





## eCapabilities Team: Lilly Web Monitor

- Monitors over 500 Lilly external websites
- Gets status (Live/Dead/Re-Directed)
- Creates basic stats on this information

Method:

1. KNIME workflow auto run by server
2. Ping each website and get status response code (i.e. 404, 200)
3. Manipulated into region based data
4. Write to DB for live display by further tool

DOMAIN	LOGGED	DEAD/RE-DIRECTED	STATUS
www.lilly.com	1	0	200
www.lilly.com	1	0	200
www.lilly.com	1	0	200
www.lilly.com	1	0	200
www.lilly.com	1	0	200
www.lilly.com	1	0	200
www.lilly.com	1	0	200
www.lilly.com	1	0	200
www.lilly.com	1	0	200
www.lilly.com	1	0	200



# Stable

Reliable,  
Reproducible,  
Transparent for  
development  
work

## SHIPPING DATA VIA WEBPAGE

1. Research IT co-develop/support workflows with GSM in development environment

Global Sample Management

3. Contractor needing limited Inventory data e.g.: shipping codes and sample info but no structures - retrieved portal

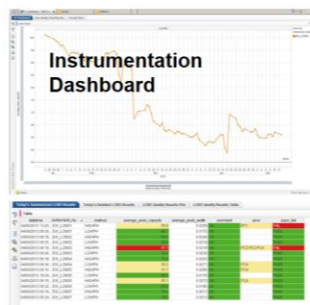
1. **Analytical Technologies** co-developed workflows with Research IT to maintain a DB of instrumentation diagnostics



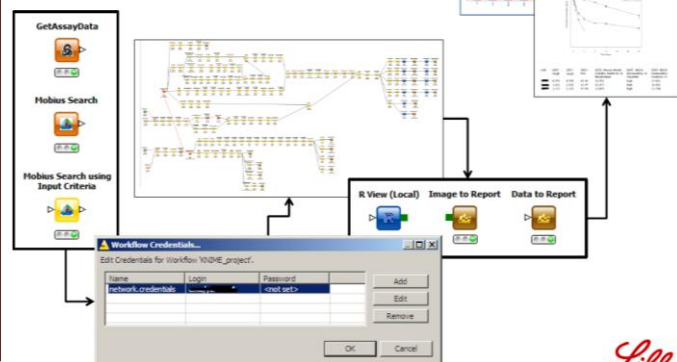
2. KNIME Server runs workflow as nightly scheduled task



3. Spotfire driven webpage tracks all AT Labs instrumentation & highlights issues (e.g.: column degradation)



Working with **ADME group** to web-enable complex workflows merging in-vitro, in-vivo and in-silico data to produce rich reports...  
...All with correct data security model (authorise data at source)



# Secure

Providing the  
right data at the  
right time to the  
right people

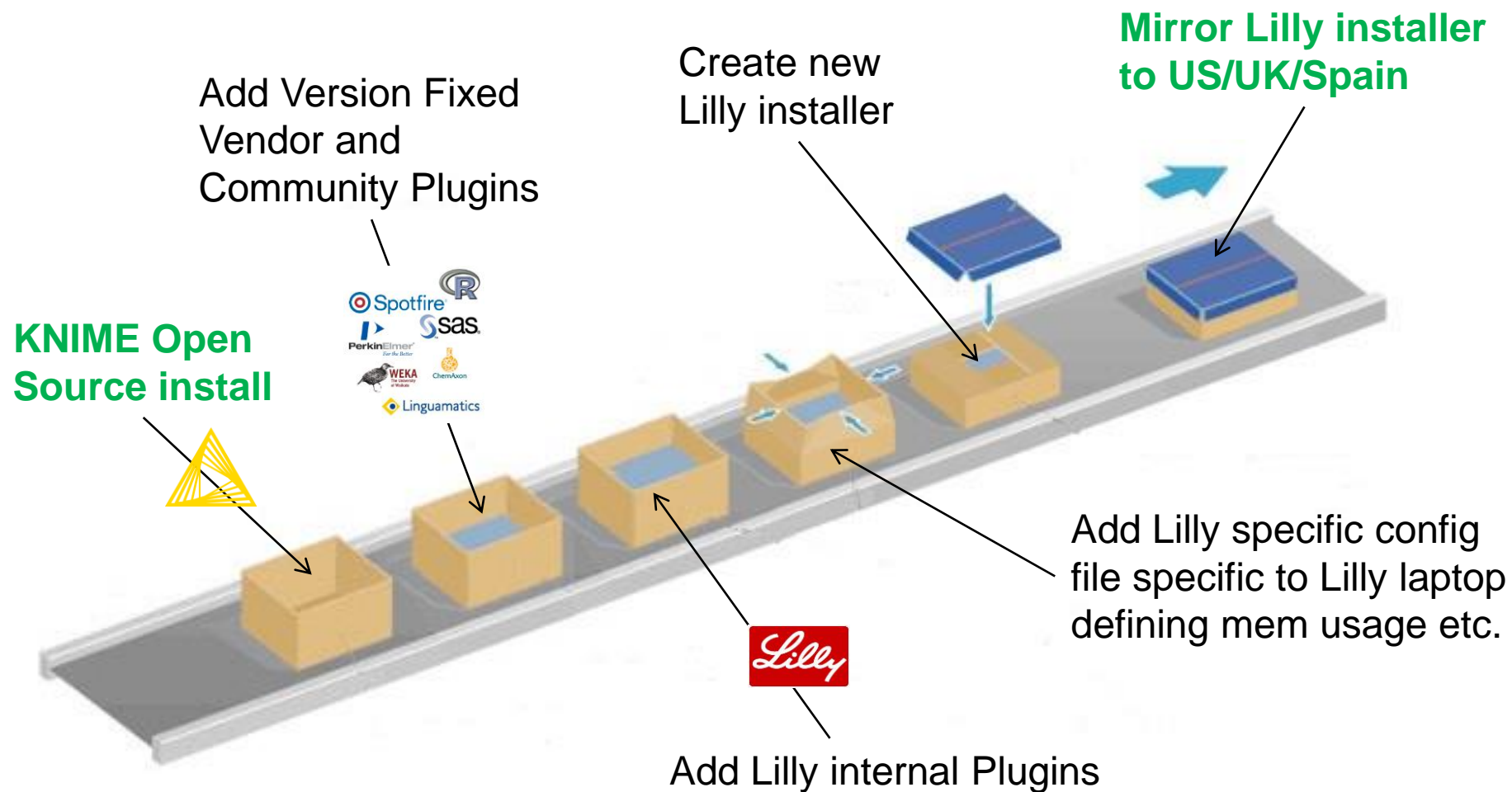
Work by: Stuart Morton / Tom Wilkin

Answers That Matter.

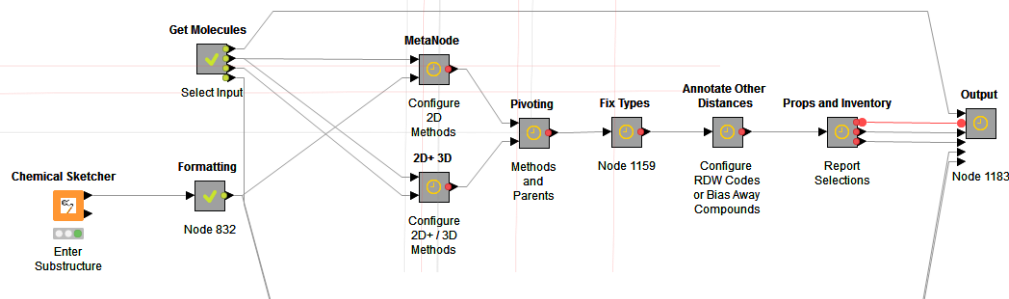
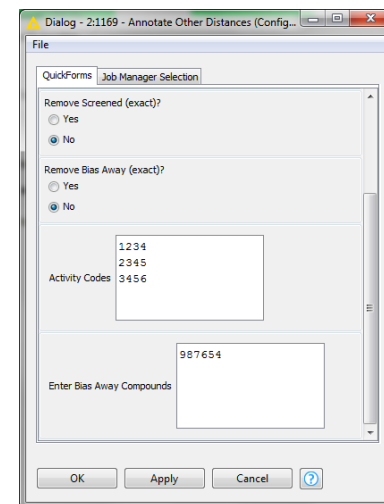
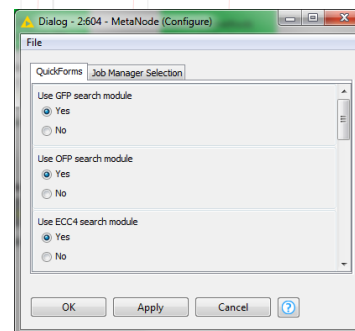
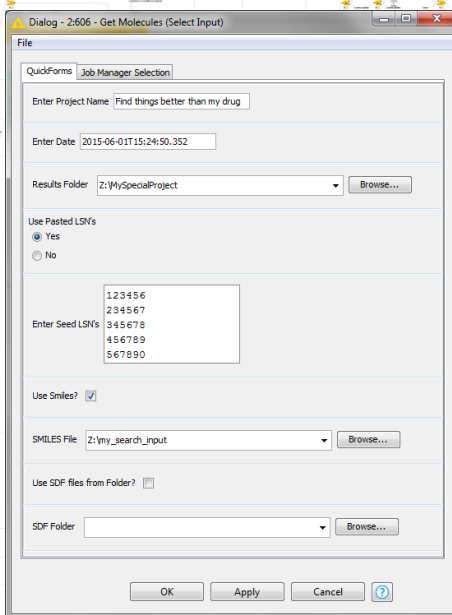
## Reshaping Data: Example Production Capabilities



- Rapid / 'Agile' development
- 1 Production KNIME version per year  
(2.6 / 2.9 / 2.12 / 3.3)
- >130 internal nodes
- 340 Features/Bugs in 2016 (heavily focused on data and data security – NTLM/OAuth)
- 6 internal updates



- 100's of nodes for data manipulation
- Data augmentation using calls to 'best of breed' cluster side algorithmic search methods
- Encapsulation of best analytics practice for problem x, towards automated reasoning

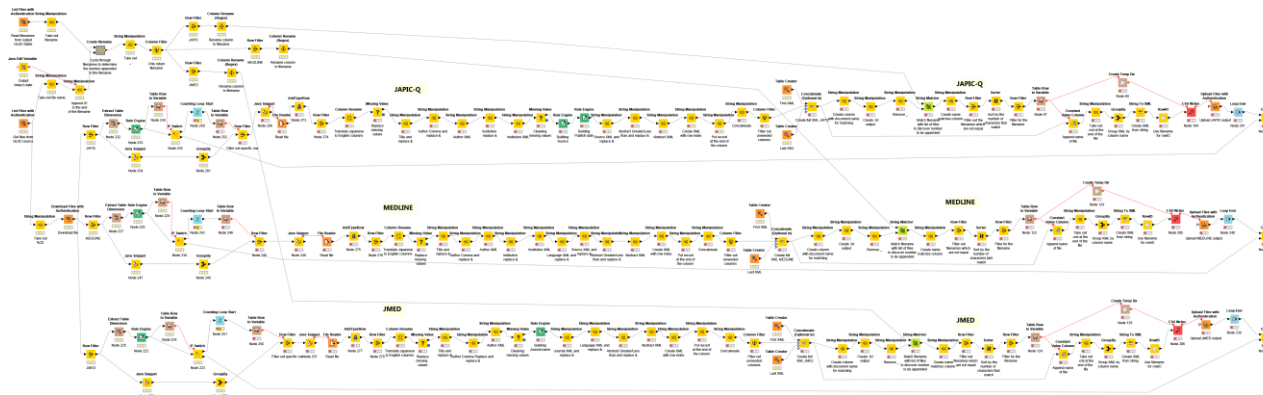




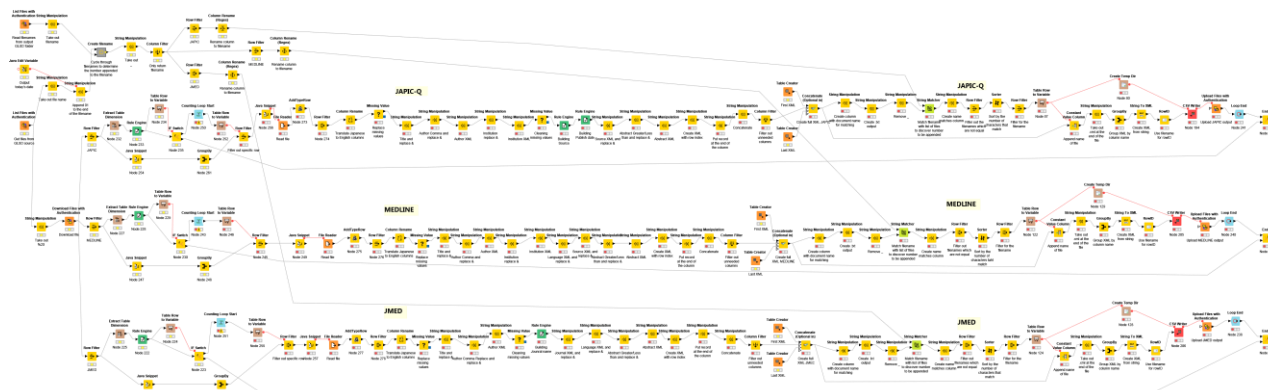
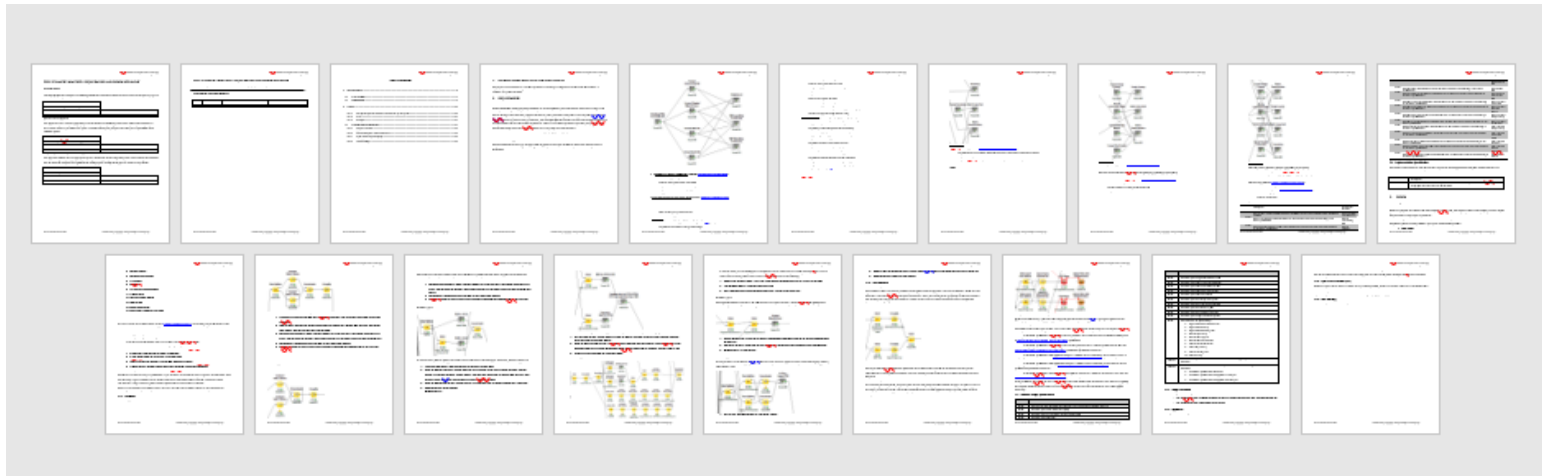
# “Here be dragons”

- Download e.g.: Medline, JAPIC-Q, JMed
- Clean, De-duplicate, Augment, Join
- Text Mine (Cogito for taxonomies and ontologies)
- Select, Alert (Email)

...simple so far



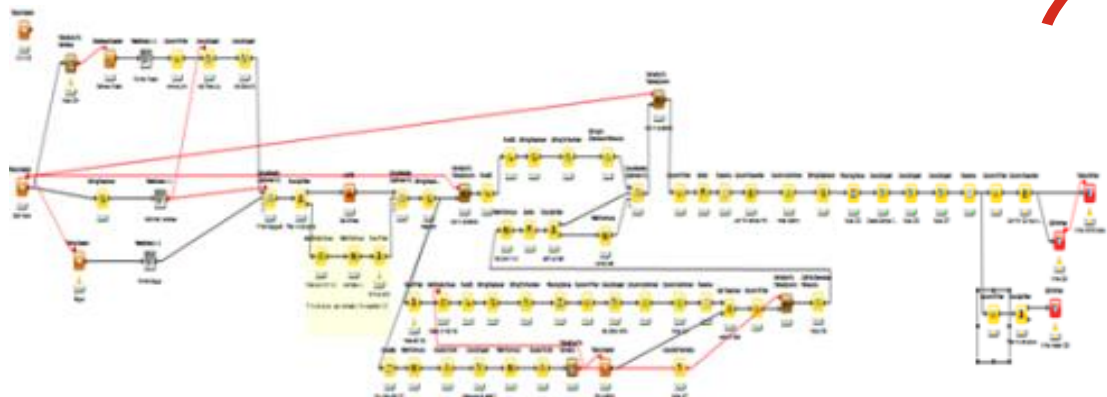
# “Here be dragons” GxP!



## Specific

The right  
analytics for my  
specific  
business  
problem

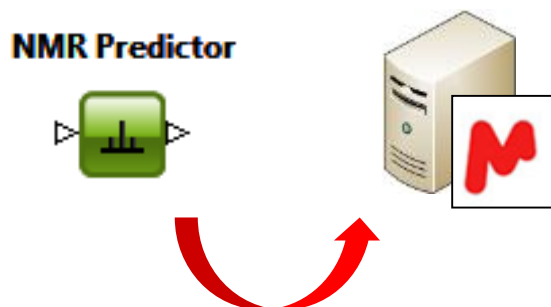




Lilly Screening Collection representing all compounds made internally or purchased for projects. Routinely used to screen new disease targets for possible starting points.

1. Extract data from multiple DB's to prepare instrument run files
2. Acquire data from instrument
3. Analyse using call to external tool server MestreLabs
4. Check results against criteria to score Pass/Fail/Review
5. Publish results to the Corporate DB

Multi year project to retrospectively check quality of all compounds run in months using KNIME



Complex and specific scientific needs only available from one vendor integrated through dedicated server build, serving predictions to KNIME via webserver API (SOA)








KNIME now analyses all new compound registrations daily (scheduled server task) to check (a) Analytical data is available and (b) is consistent with the registered structure and (c) emails submitter on Fail



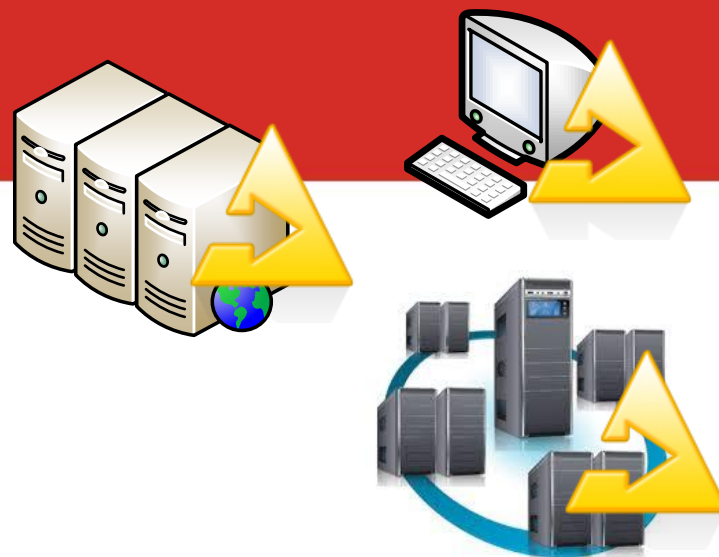




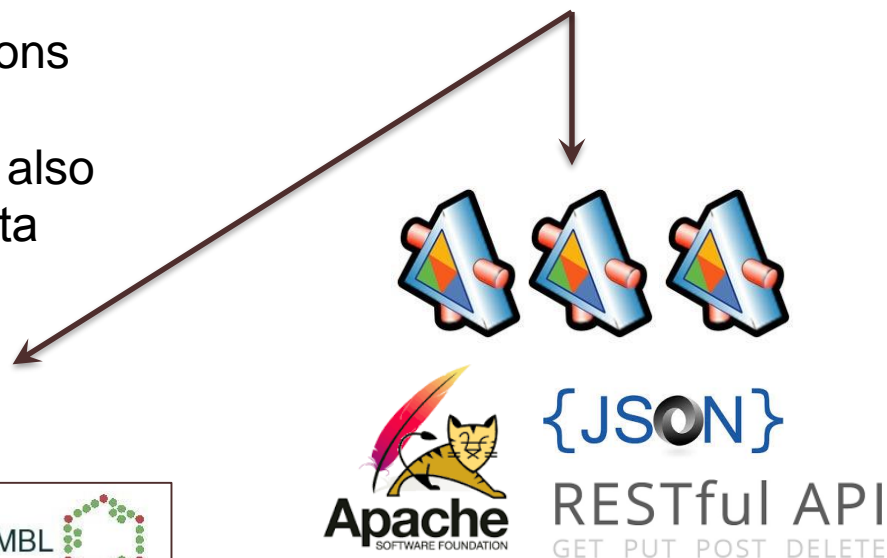
Reshaping the Architecture: >140  
RESTful Webservice API endpoints  
deployed in 2-3 years primarily for data  
retrieval and augmentation by KNIME

- ▲  Lilly Data Access
  - ▷  Assays
  - ▷  Mobius
  - ▷  Other
  - ▷  Samples
  - ▷  Structures
  - ▷  Curve Service



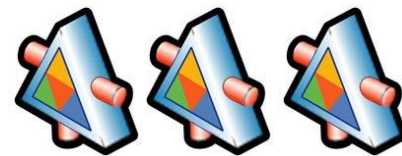
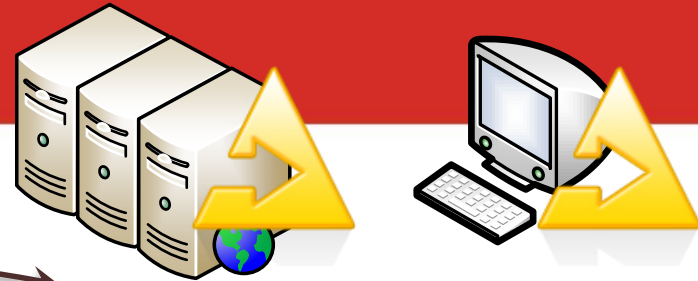


Many legacy, new and vendor applications deployed internally with webservice endpoints for KNIME integration. Team also built many API's (& nodes) for many data sources





- Multiple other internal systems now utilising KNIME, but also independently accessing the SOA layer for programmatic data access



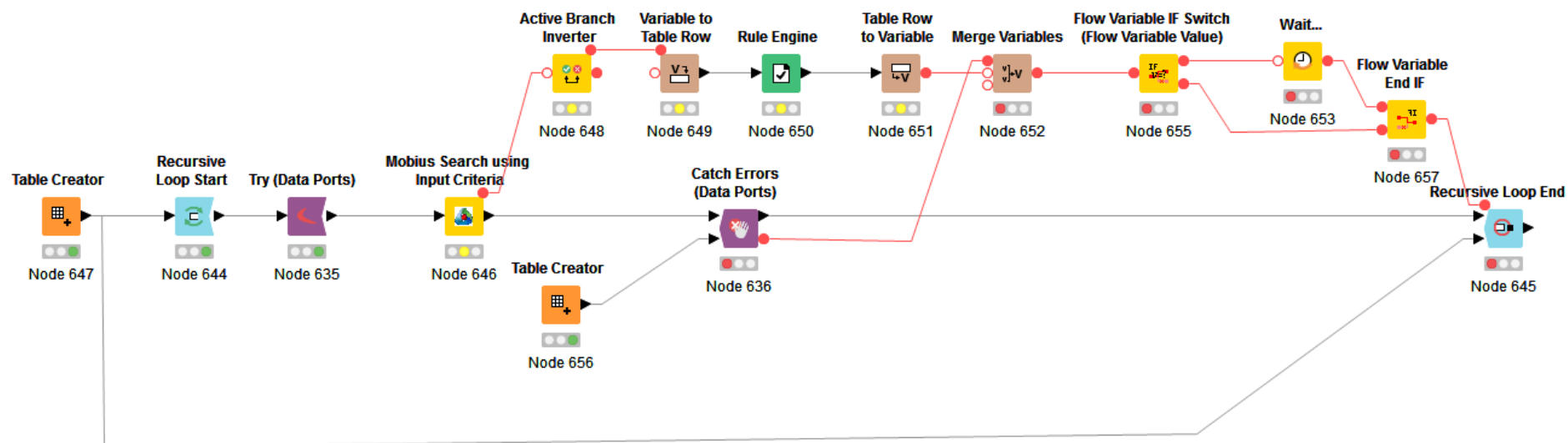
## Integrating via SOA:

- New KNIME versions
- Internal feature requests (tools/data)
- Data security
- Data-load/Volume
- Streaming

*understood &  
well managed*

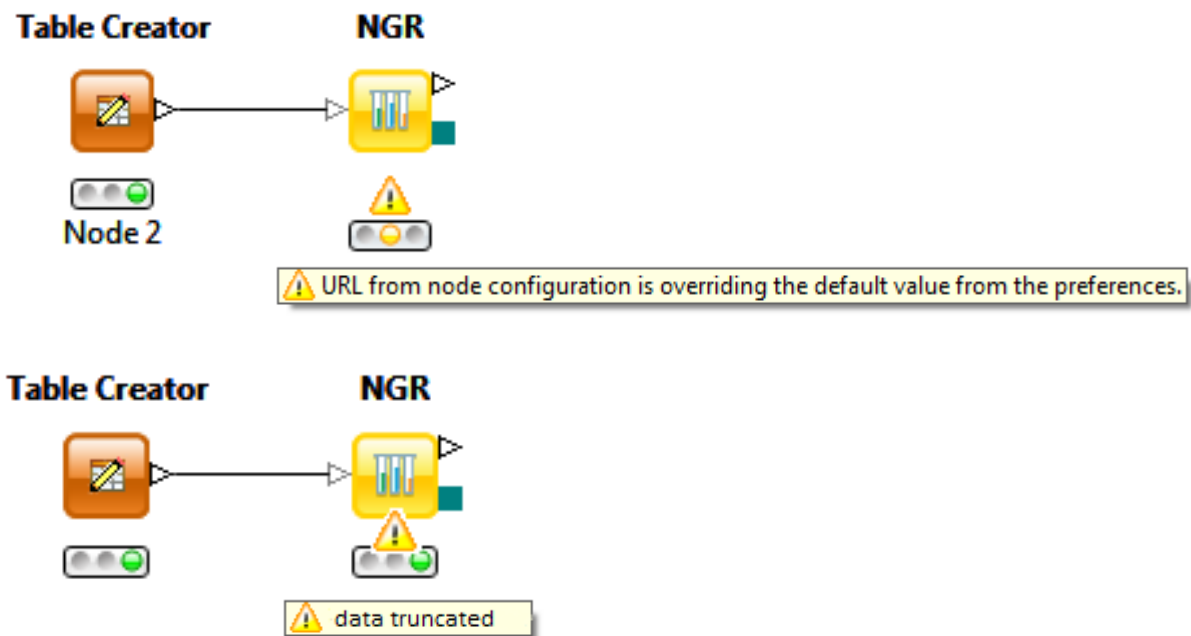
*Challenges remain*

## Timeouts and retries

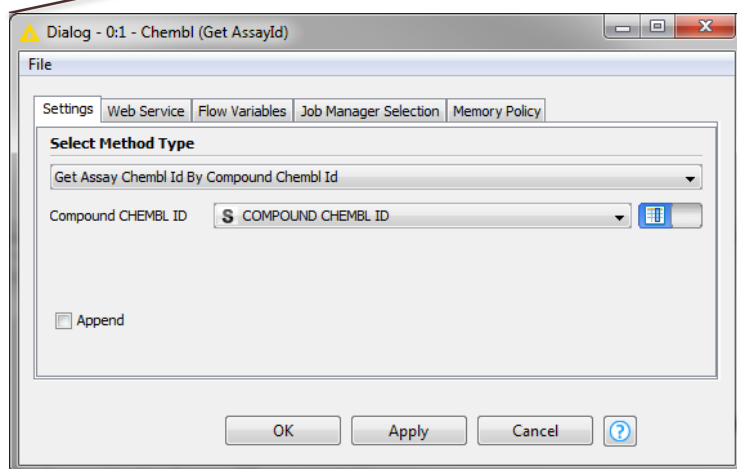
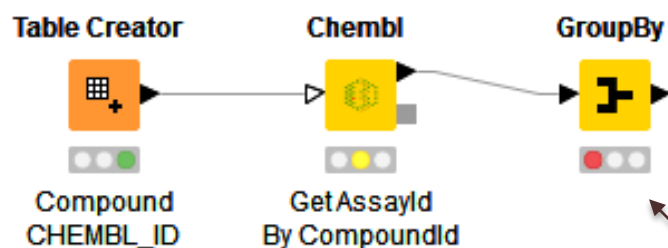




## URL Over-ride and data truncation:

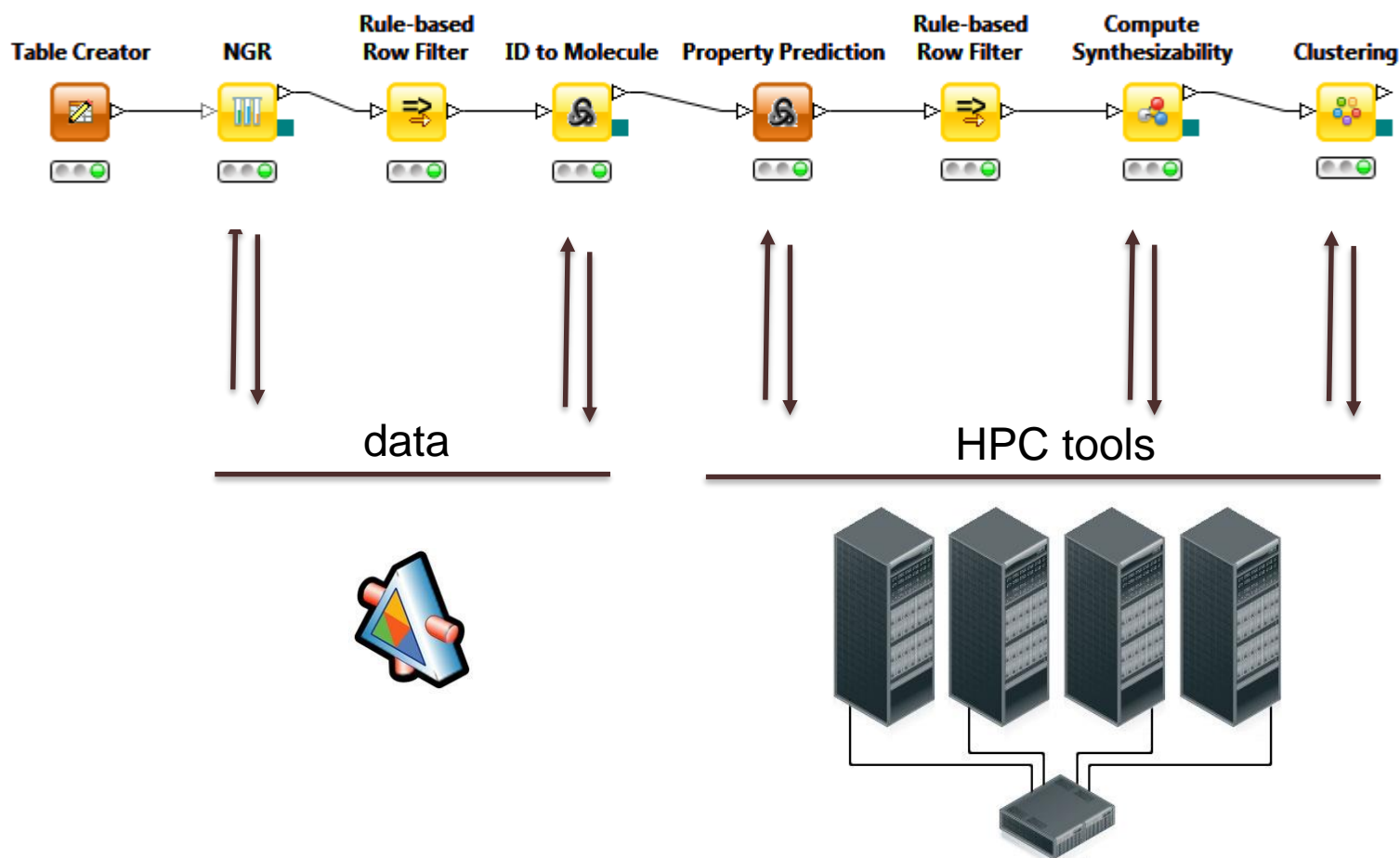


## Trade offs: Downstream configure...

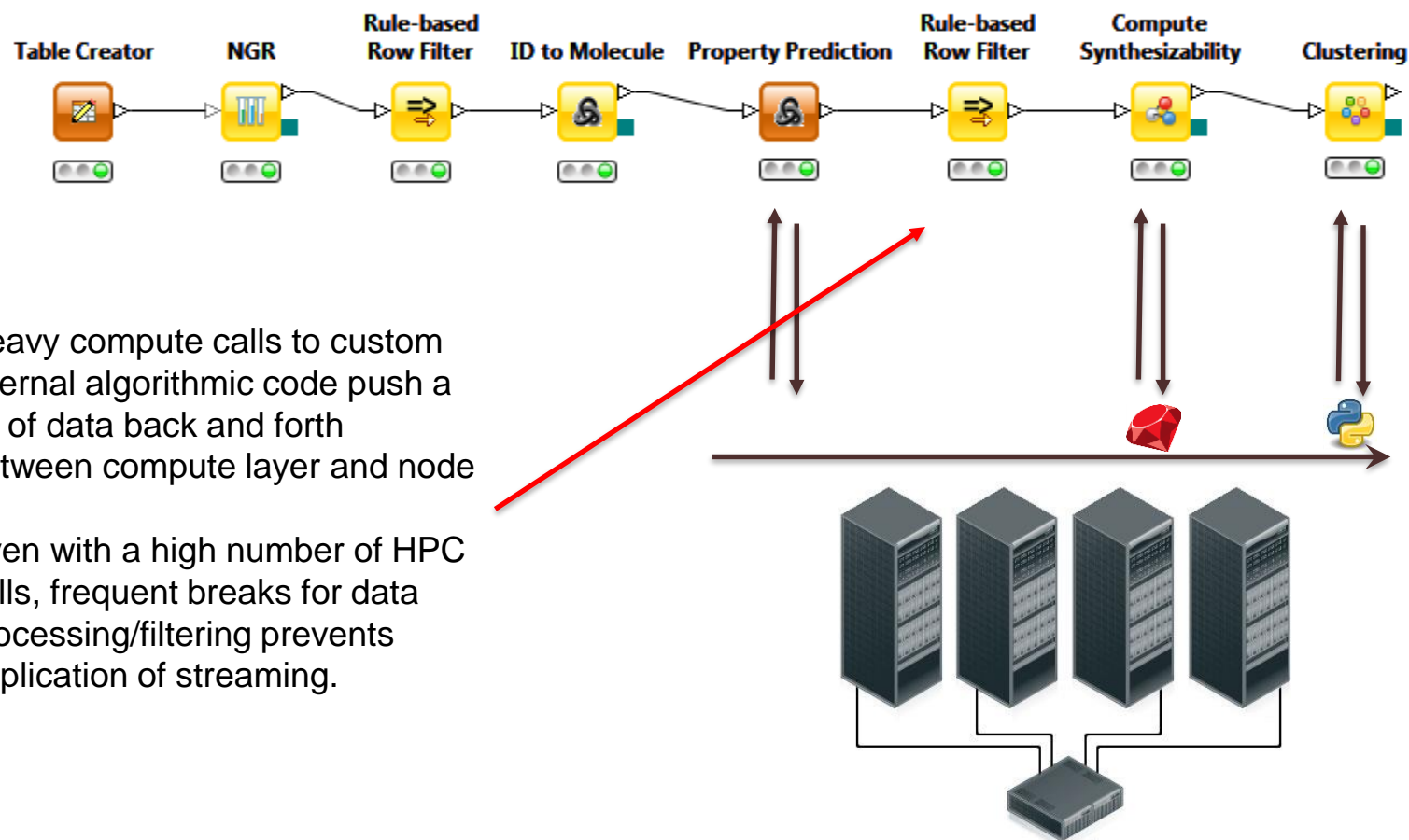


- Internal Mirror of many external data sources for secure querying
- Between version API changes (schema upgrades) break nodes if return columns are hard coded
- Alternate flexible column specification on return breaks downstream configure

Bottlenecks and frequent https calls:



## How to break streaming:



Heavy compute calls to custom internal algorithmic code push a lot of data back and forth between compute layer and node

Even with a high number of HPC calls, frequent breaks for data processing/filtering prevents application of streaming.

## Scalable

Can it handle the  
volume and variety  
of our data

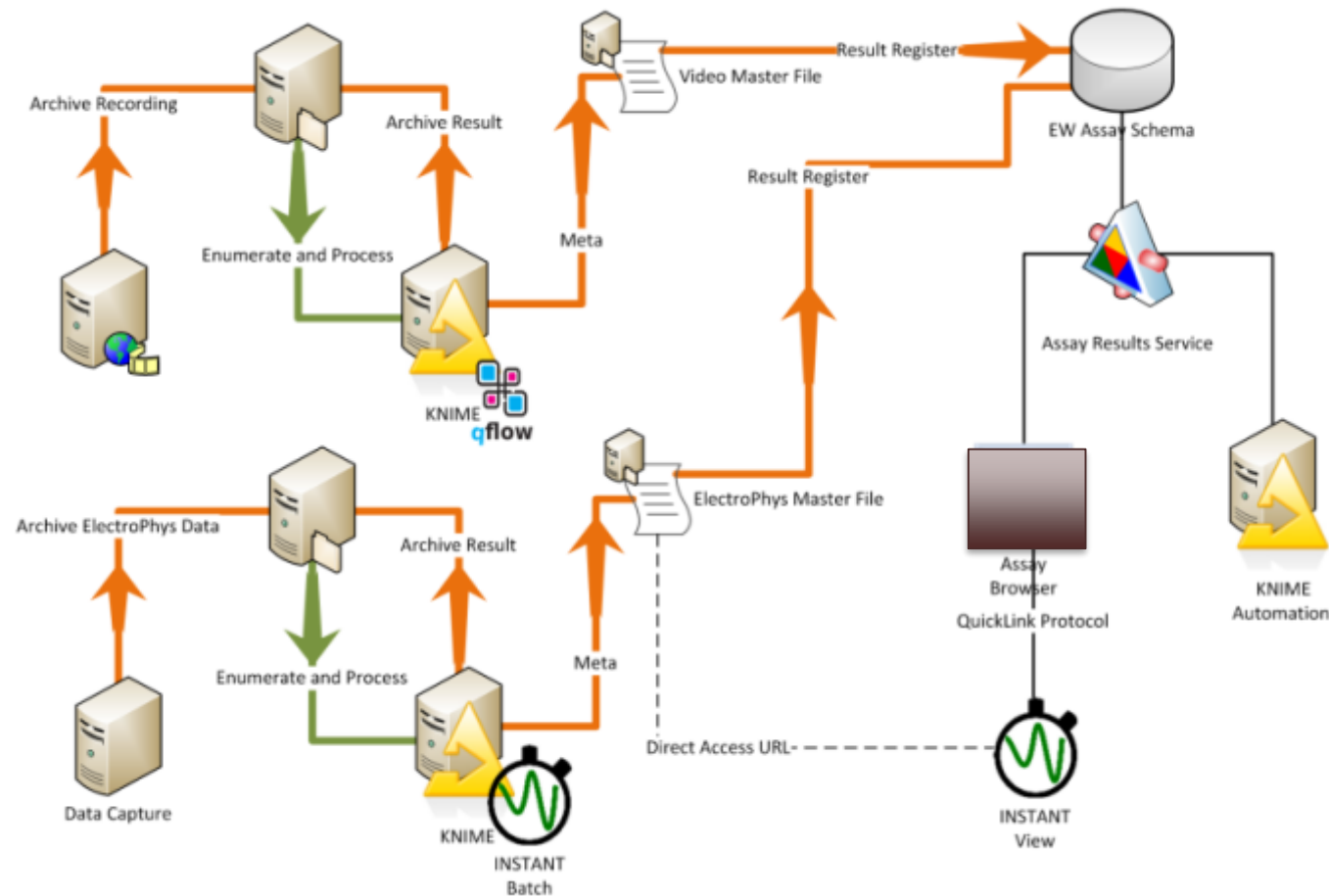




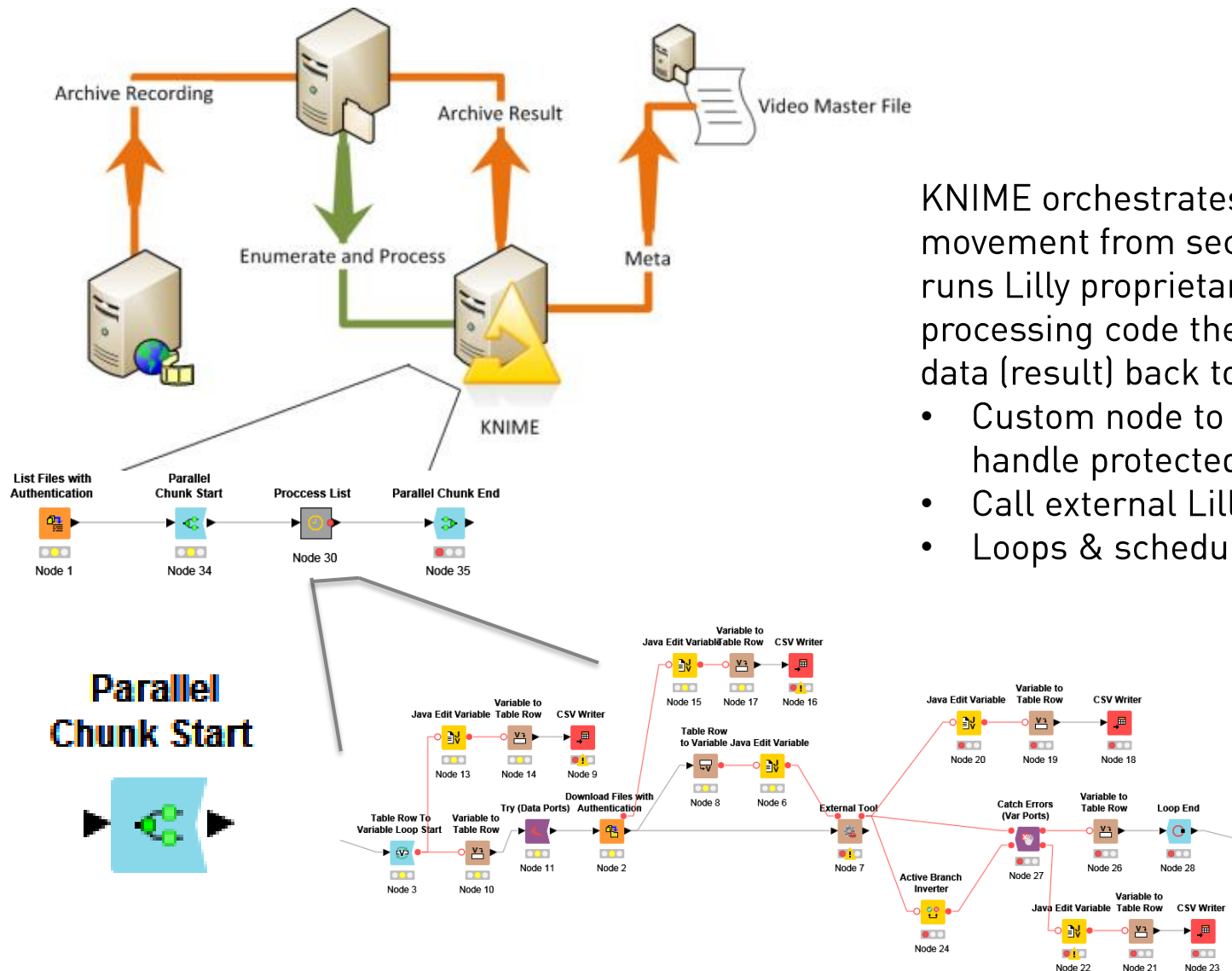
# Video feed Meta data for KNIME Analytics

KNIME processes and integrates additional data streams alongside video feed

User front end (INSTANT) provides interactive analysis of meta data and ad-hoc analytics via KNIME automation



# Scaling out Orchestration with Parallel Chunk



KNIME orchestrates file movement from secure file store, runs Lilly proprietary signal processing code then writes meta data (result) back to disk or DB:

- Custom node to correctly handle protected file shares
- Call external Lilly code
- Loops & schedule to automate



**Eli Lilly & Company** 

@LillyPad



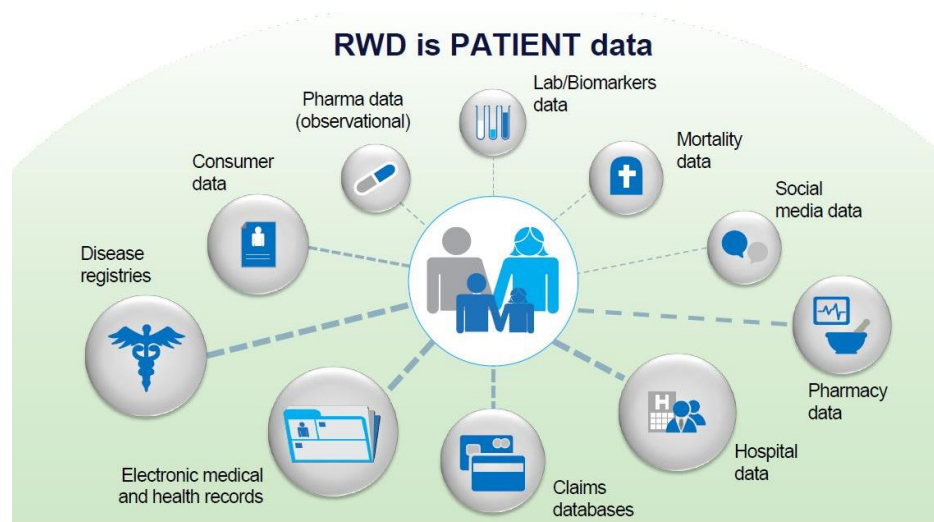
Follow



Real-world evidence helps us discern all the ways a medicine can be used. [elil.ly/534g](https://elil.ly/534g)  
#pharma

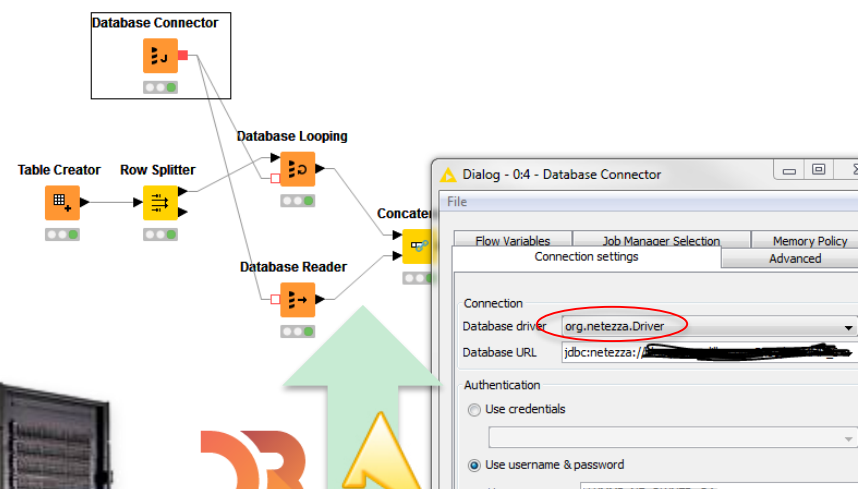
5:27 AM - 25 Mar 2016





DOI: 10.7365 / JHPOR.2015.2.4

*How is an existing medicine X and treatment Y working in the health care system?*



*Querying Netezza (IBM Pure Data) Store: Metrics, Sentiment, Outcome...*

**Research:**

Dave Evans  
Lewis Vidler  
David Thorner  
Jason Ochoada  
Christos Nicolau  
Beth Wright  
Gary Sharman  
Sylvain Demanze  
Natalie Franklin  
Steve Green

**Research IT:**

Luke Bullard  
Thomas Wilkin  
Nathan Roberts

Stuart Morton  
Malika Mahoui  
Jeramy Brewster

**Other API & System Owners:**

Hongzhou Zhang  
Jim Hughes  
Christos Nicolaou  
James Rimell

... and the many other users and developers who broke things, fixed things, and did neat things with KNIME