

Primetime for KNIME:

Towards an Integrated Analysis and Visualization
Environment for RNAi Screening Data

F. Oliver Gathmann, Ph. D.
Director IT, Cenix BioScience

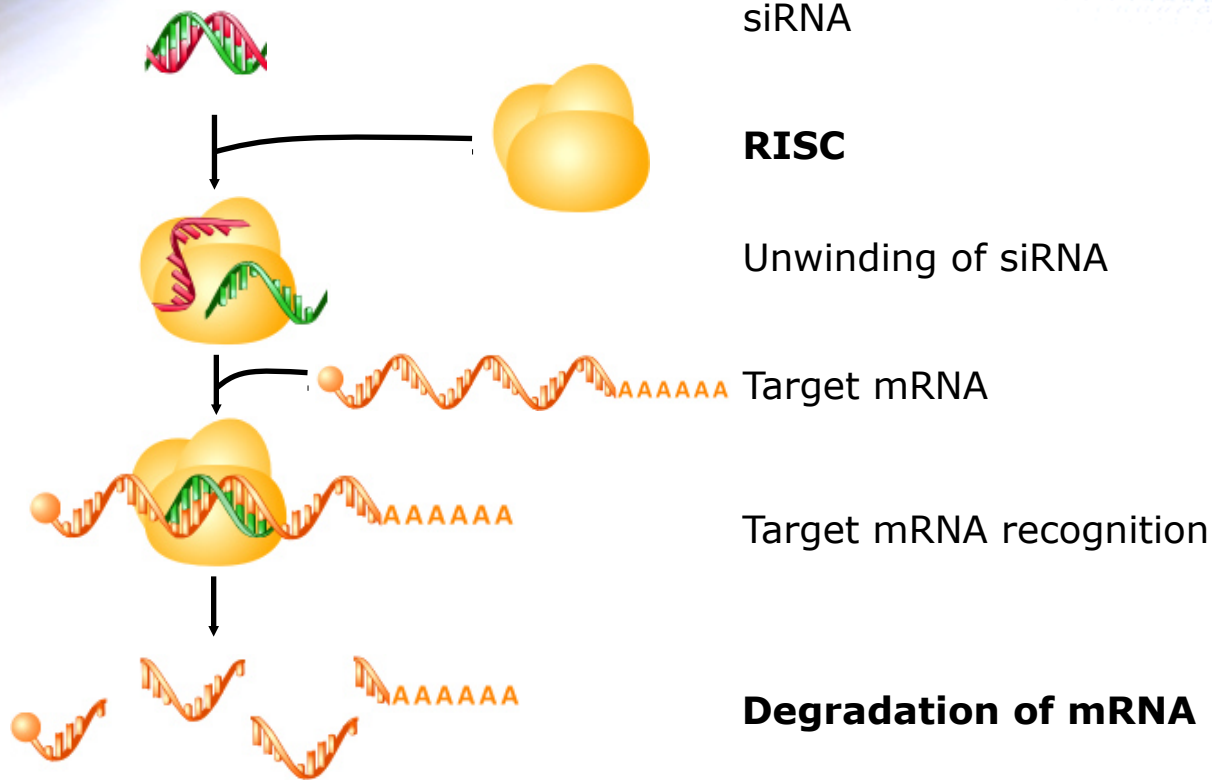
Presentation for:
KNIME User Group Meeting 2011

Zürich, March 3rd 2011

Overview

- Explain “RNAi Screening”
- IT infrastructure for HT-HCS
(High-Throughput, High-Content Screening)
- Workflow software evolution at Cenix: past, present, and future

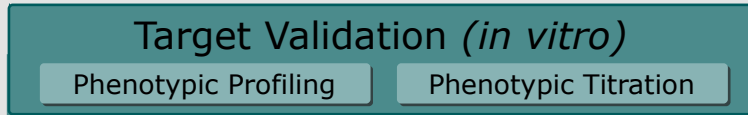
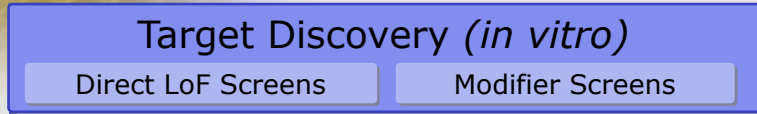
How RNAi works



- **First Take Home Message:**

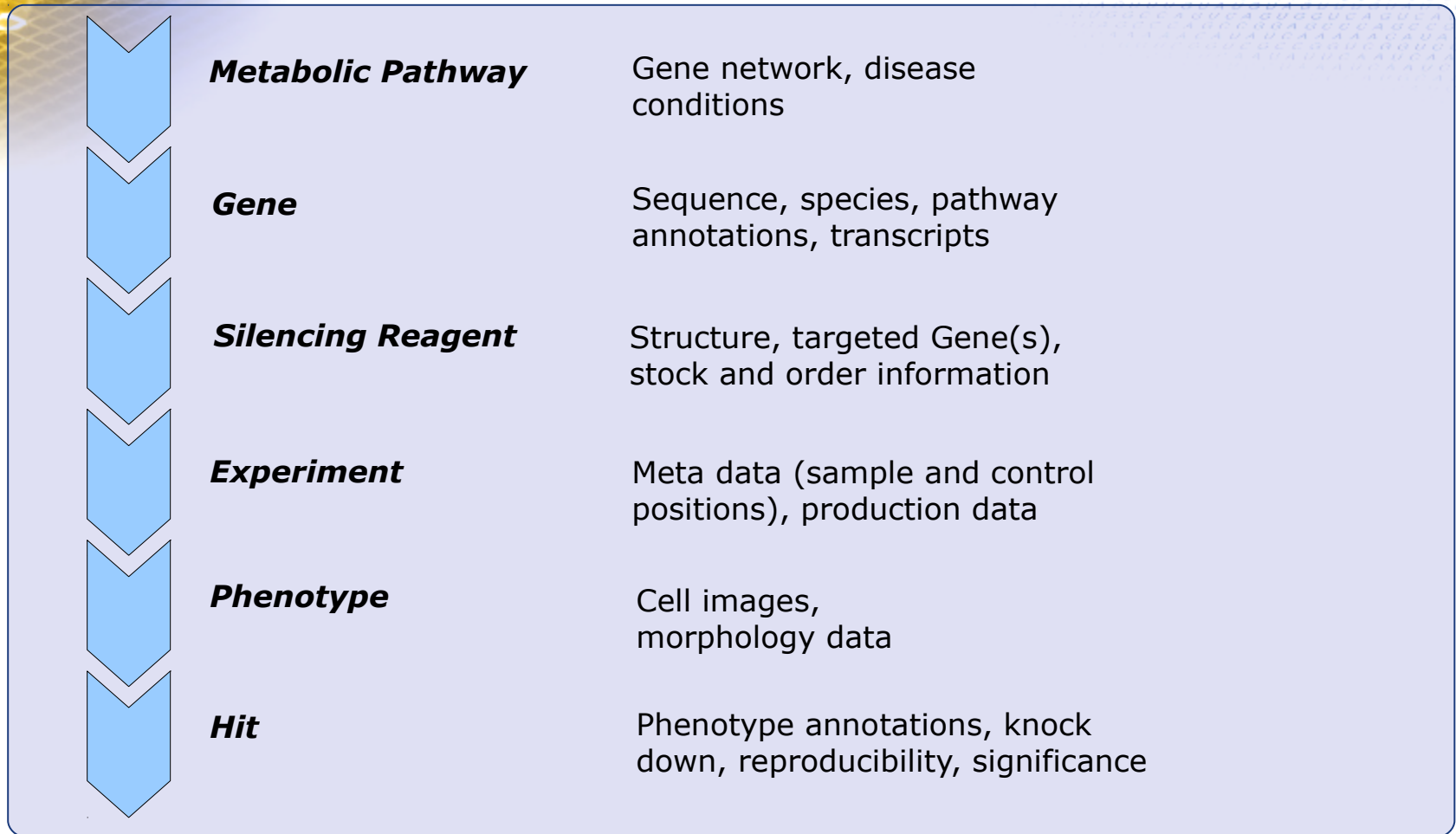
RNAi allows you to investigate the function of genes by knocking them down selectively

The Drug Discovery Pipeline



- **Second Take Home Message:**
"Early In The Drug Discovery Pipeline" means high-throughput and lots of data

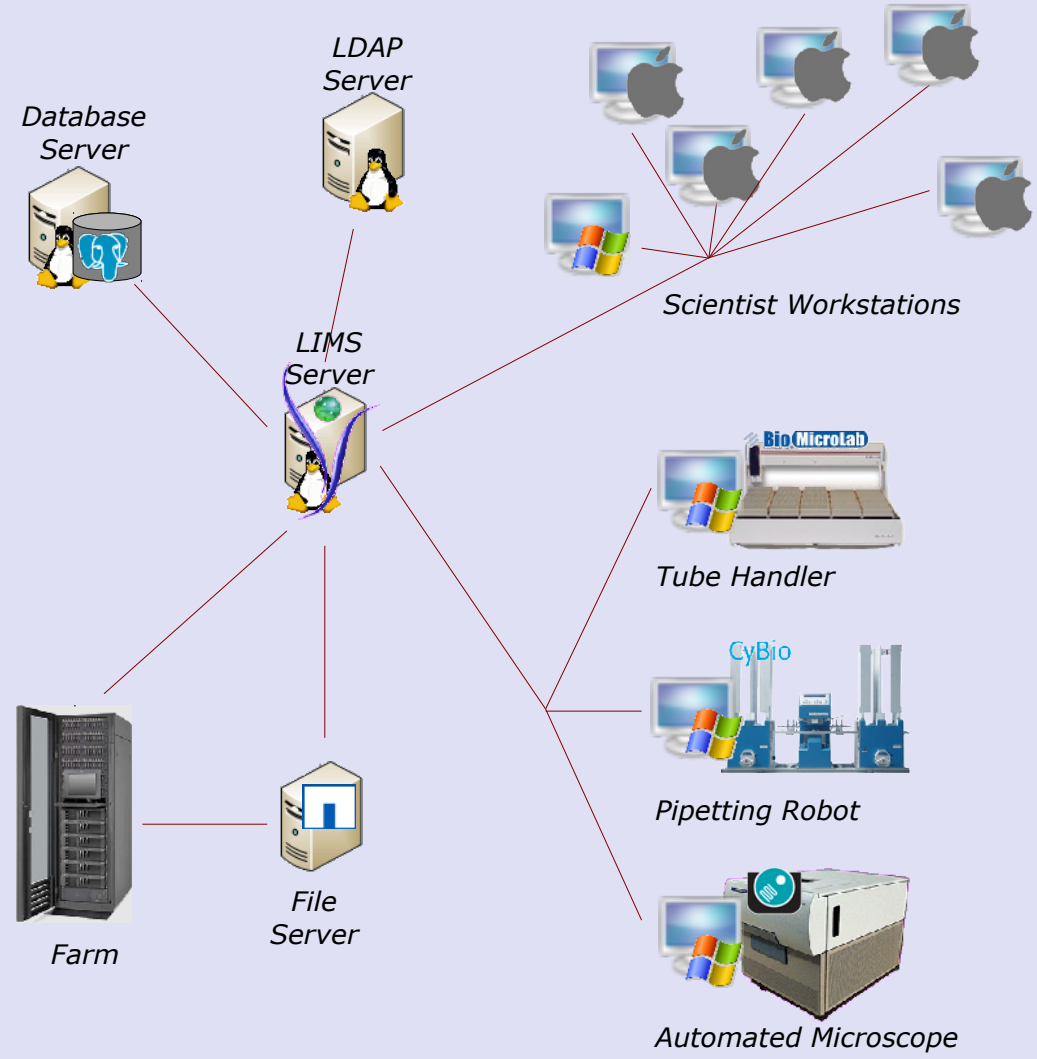
Information Layers



- **Last Take Home Message:**
"High-Content" means complex data structures

IT Infrastructure for HT-HCS

Cenix

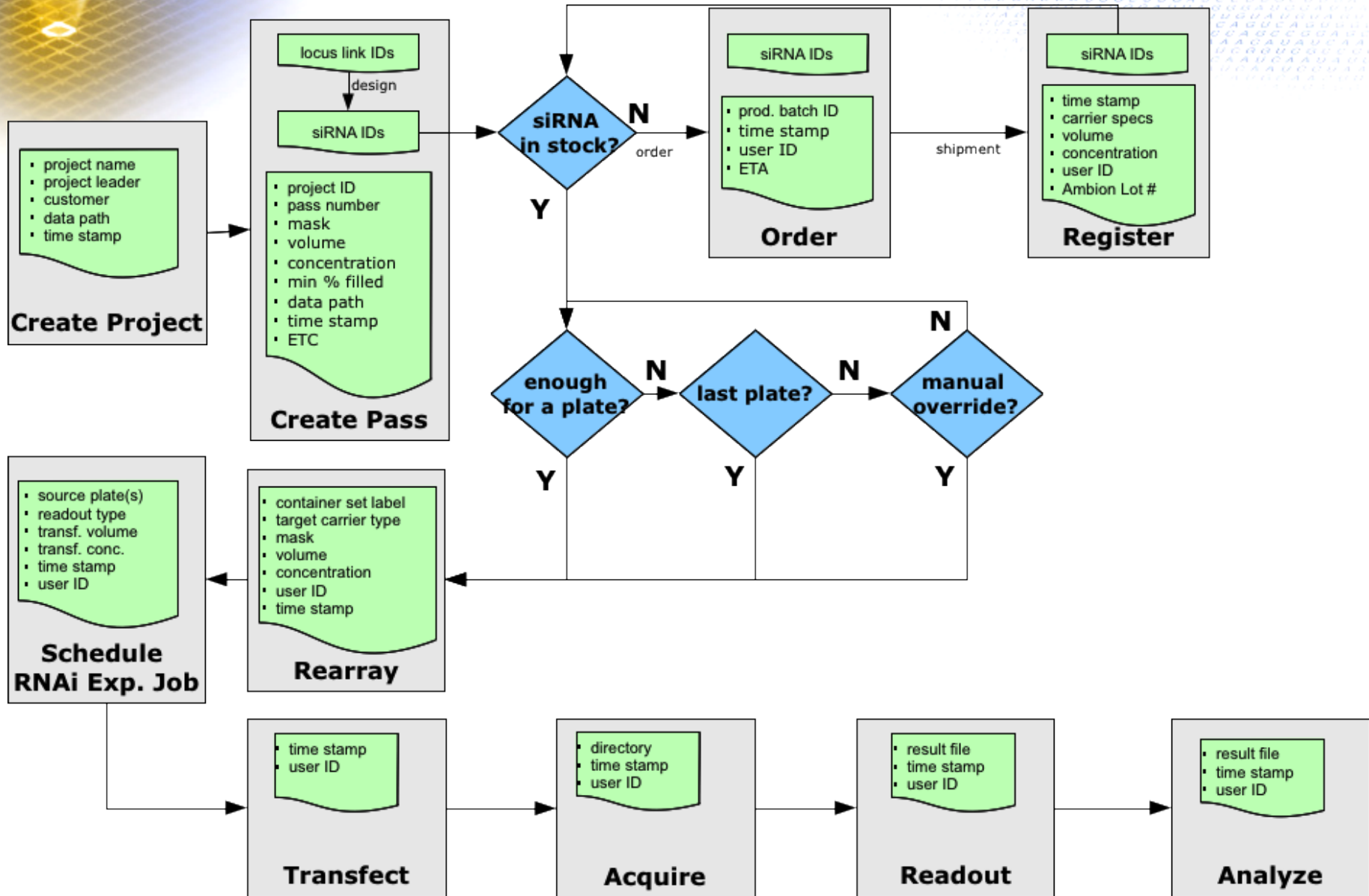


Terminology: "Workflows"

"Process-centric Workflows" vs. "Data-centric Workflows"

- Process-centric: mapping a work process in the physical world; focused on data *acquisition*
- Data-centric: mapping an algorithm; focused on data *processing*
- Not always clear-cut, but still useful distinction

Primordial Process Workflows: Design



Primordial Process Workflows: Implementation

Cenix Lab Management

About | Create And Schedule | Monitor And Control | Analyse And Query

ID	Job Name	Job Type	Submitted by	Status	Start Time	Project	Experiment
11	Mitotic Index Assay	AutoScope RNAi Experiment	M.Hannus	QUEUED	-	21: Oncology Screen IV	68: Follow-up

Workflow Diagram:

- Step 1: RNA Transfection (Icon: Microplate)
- Step 2: Data Readout (Icon: Microplate with pipette)
- Step 3: Image Processing (Icon: Microscope)

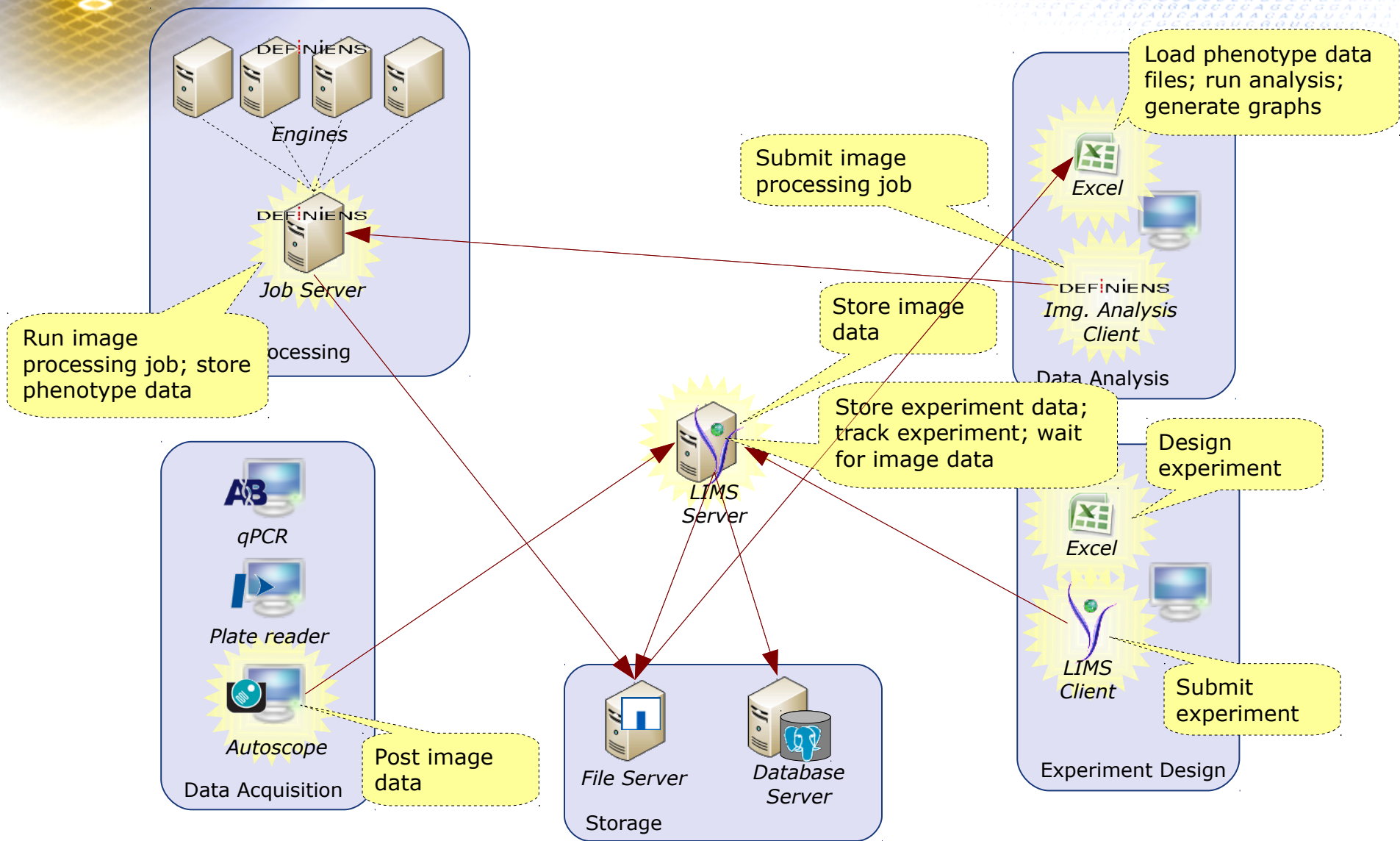
Callout Boxes:

- seamless integration of robotics** (Image: Laboratory robot)
- automated capture of experiment data** (Image: Microscopy image of cells)
- user-driven data analysis** (Image: Heatmap and bar chart)

Data Analysis Workflows: Excel

- In the beginning, there was Excel.
 - + Advantages:
 - Ubiquitous and easy to use
 - Full flexibility for the end user (in theory, anyways)
 - Disadvantages:
 - Hard to debug
 - Nightmarish version control
 - Slow and cumbersome

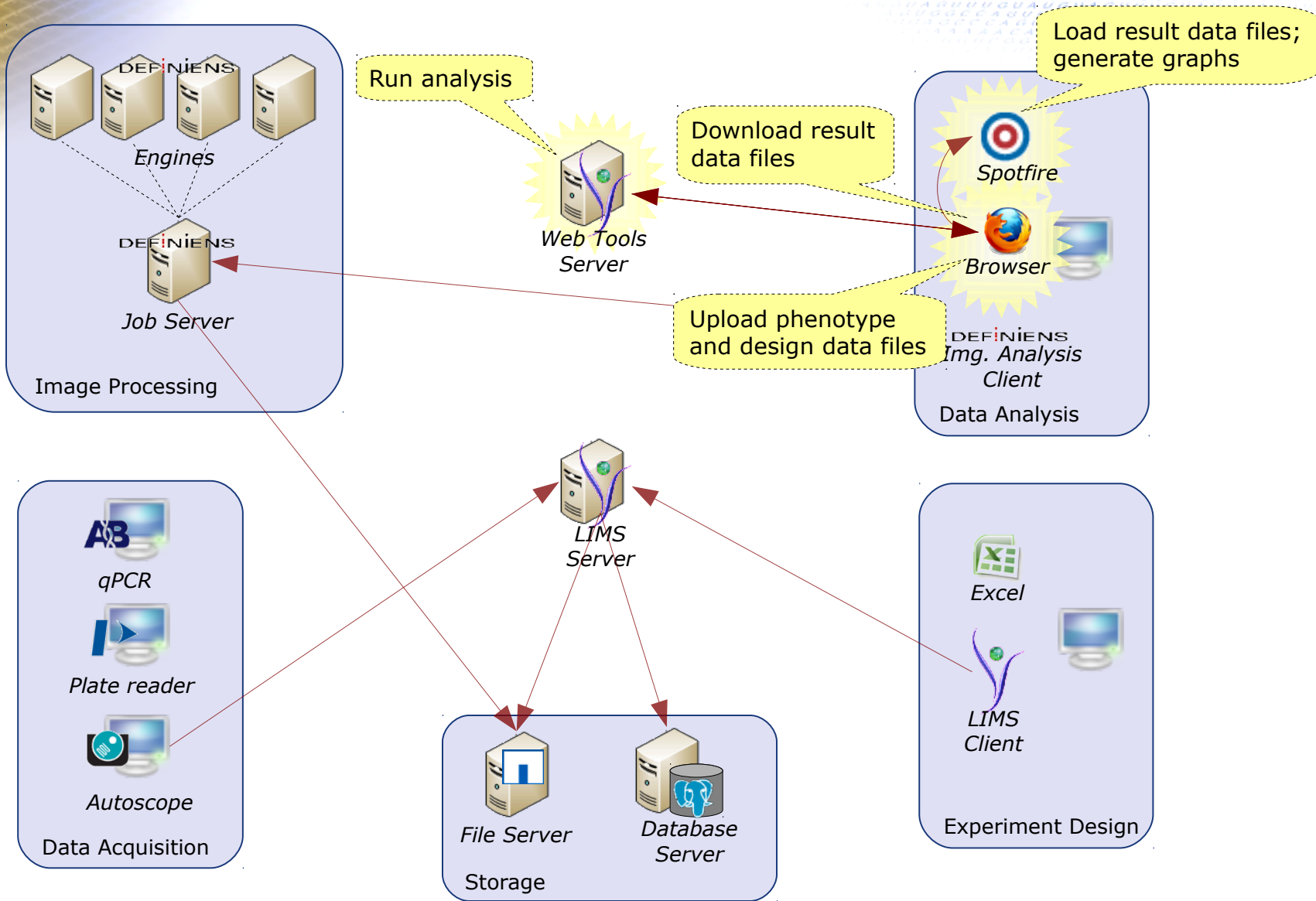
Data Analysis Workflows: Excel



Data Analysis Workflows: Web Tools

- Next: Web tools with tabular data as input and output.
 - + Advantages:
 - Encapsulation of complex functionality
 - Centralized administration
 - Executed on server
 - Disadvantages:
 - Low flexibility
 - Frugal web interface

Data Analysis Workflows: Web Tools

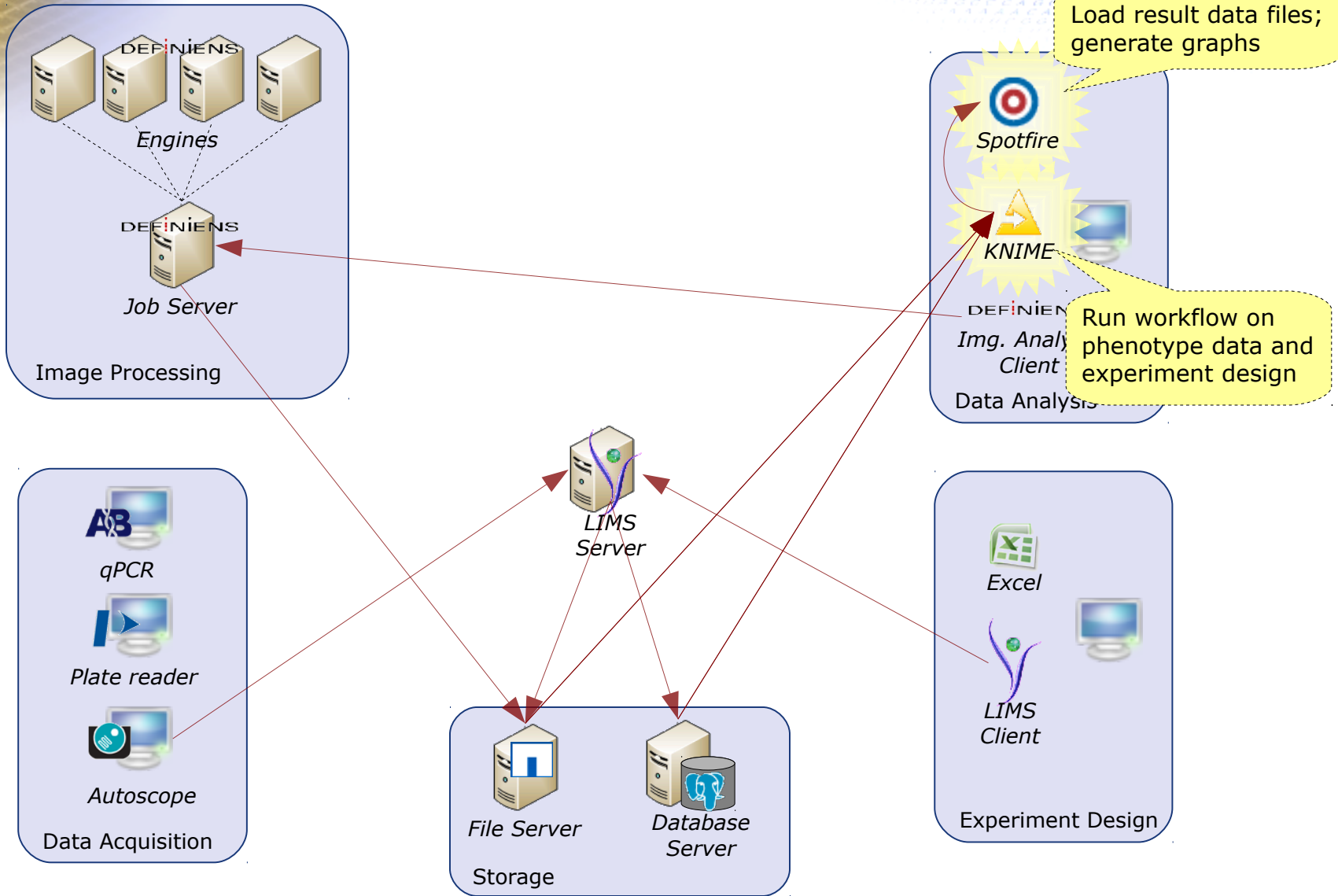


Data Analysis Workflows: KNIME!

- **KNIME: A giant leap forward**
 - Flexible and easy to use and yet robust, scalable, performant and extensible!
- **Current KNIME infrastructure:**
 - Centrally administered Windows and Mac installations, configured to point to a user-specific workspace on the file server
 - Workflow curation policy: Versioned reference workflows for each project, owned by power users
 - Experiment meta data provided through database nodes, raw data through files
 - Complex statistics implemented with (remote) R scripting nodes



Data Analysis Workflows: KNIME!

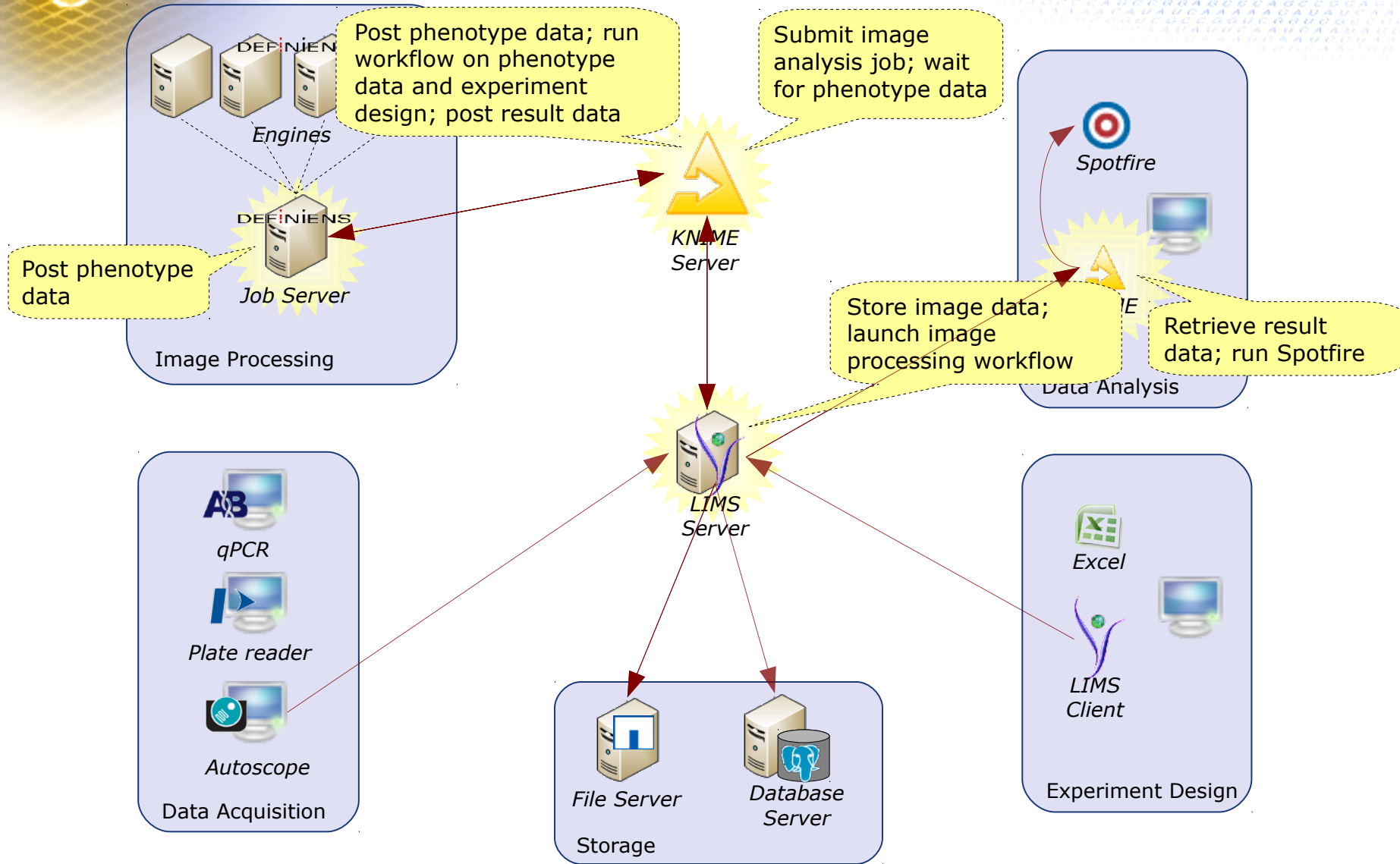


Primetime: Requirements

- Streamlining the Screening Pipeline
 - Analysis has become the bottleneck: Potential for 10-20 % increase in overall throughput
- Even “Higher” Content:
 - More parameters using advanced analysis methods
 - Single object rather than population data
 - Integrate gene annotations and pathway data
- Enable customers to explore and (re-)analyze delivered data sets
 - Selecting/weighing parameters
 - Tight integration with Spotfire, including raw data

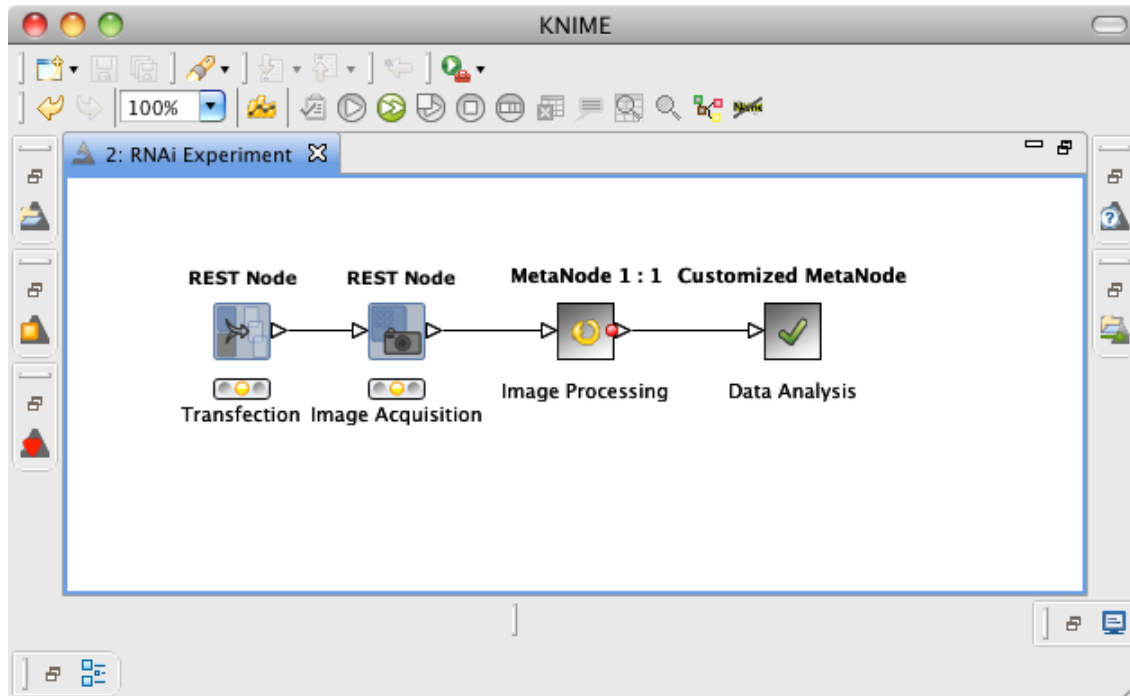
Primetime: IRIS

"Integrated computational environment for high throughput RNA Interference Screening"



Primetime: Beyond IRIS

- Use KNIME for process-centric workflows as well
- This would require
 - Standard interface to the LIMS server to drive the “business logic” (REST)
 - Easily configurable User Interfaces to parameterize processing steps (something like RGG?)



Primetime: Beyond IRIS

- KNIME “solutions”: Hide complexity of workflows by exposing only a few “knobs” to the end user
- Features:
 - Again, a User Interface generator to make it easy for non-IT power users to create new solutions
 - Ideally, a way to publish the “solution” to a server and run it remotely

Conclusions

- KNIME has quickly become an integral part of the HT-HCS screening pipeline at Cenix
- Current work on the data analysis infrastructure around KNIME is focused on tight integration with the LIMS server, with Definiens for image processing, and with Spotfire for data visualization
- Further down the road, we plan to use KNIME for all workflows at Cenix and to build pre-packaged “solutions”

Thank you! Any questions?