Analytics

Optimization of raw materials genealogy in drug manufacturing with R, Shiny and d3

08/15/2018

Tanya Cashorali
@TanyaCash21

SPEAKER PROFILE

TANYA CASHORALI

@TANYACASH21

2004 — Partners HealthCare Center for Personalized Genetic Medicine

2005 — Children's Hospital Boston

2006 — Dana-Farber Cancer Institute

2007 — GNS Healthcare; dataspora

2012 — basho

2013 — comlinkdata

2014

2015 — Biogen

Our Services

Machine Learning

Statistical modeling to continuously learn and derive insights from data

Data Visualization

Visual depiction of data through web applications and dashboards

Advanced Analytics

Predictive modeling to capture trends in data and evaluate future impacts of decisions

Data Management

Capture and storage of data for efficient use in workflows and analysis

Business Intelligence

Analysis of historical and current business operations to identify areas for improvement

Data Strategy & Enablement

Strategic assessments and training on data capture, management, analysis and visualization

Life Sciences Solutions

CASE STUDY: BIOGEN

RAW MATERIALS OPTIMIZATION

Biogen

- Pioneer in biotechnology, with a portfolio of pharmaceuticals for neurological and neurodegenerative conditions
- Uses novel science and leading-edge technologies to create, commercialize, and manufacture transformative therapies
- Ideology for incorporating innovative technologies into their operations
- Recognized an area for optimization – raw materials genealogy in drug manufacturing.

Biogen

AVONEX (interferón beta-1a)

Tecfidera (dimethyl fumarate)

TYSABRI (natalizumab)

Project Background

Functional

- Combinations of raw materials in specific sequences are used to generate final drug substances
- Even in compliance with GMP standards, major deviations can occur
- When a deviation or discrepancy is identified, it is critical to stop manufacturing of all batches that include the problematic substance
- Identifying the incident's root cause is a complex endeavor

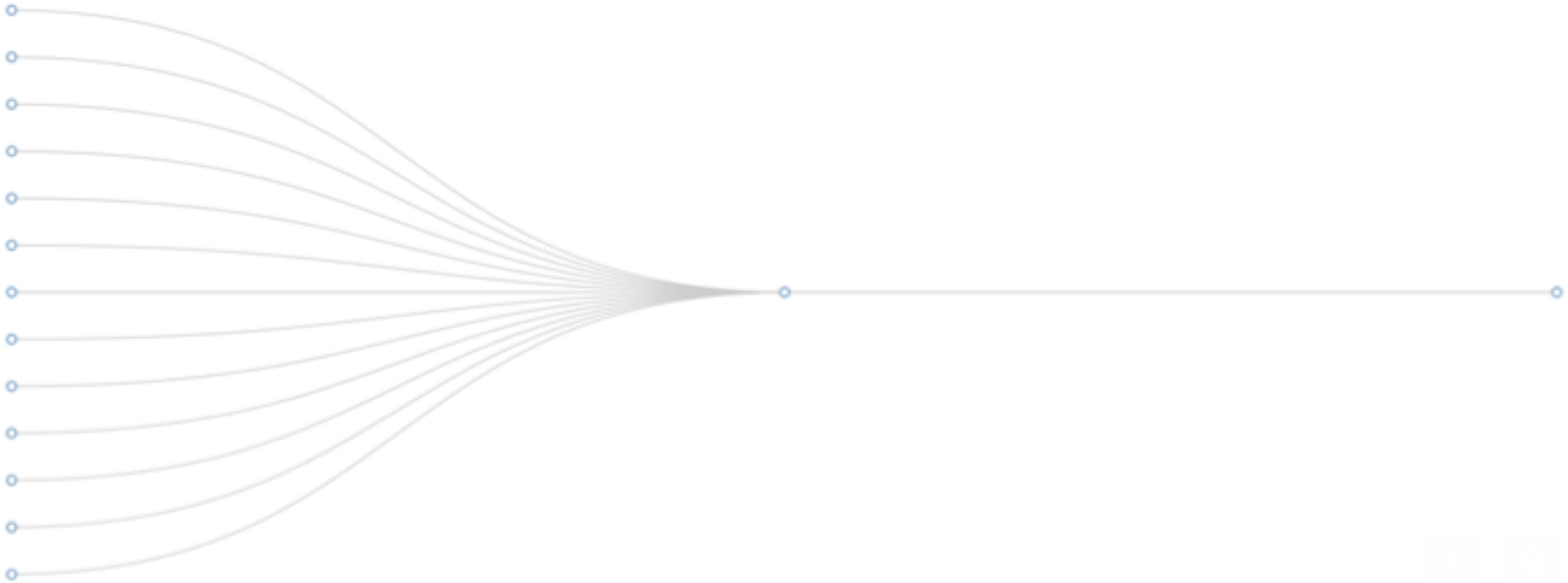
Project Background

Technical

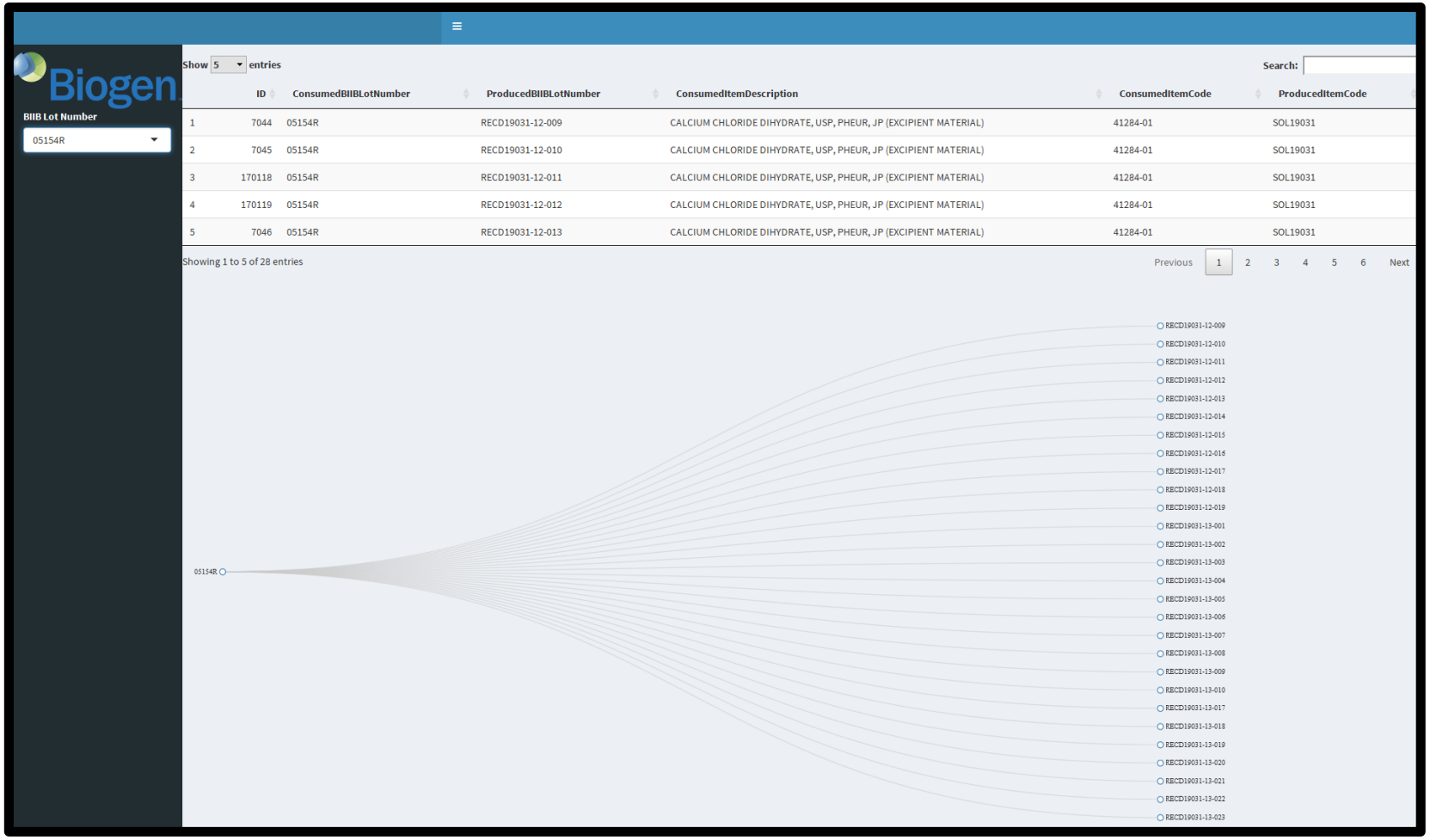
- Biogen's existing process to identify deviation root causes was manual and time consuming
- Siloed electronic systems with missing information
- Information submitted from raw materials vendors was paper-based
- Existing analysis of the information was manual, with a single user running potentially hundreds of SQL queries to identify the root cause of deviations

networkd3

```
{
    "root": "RECD19135-15-048",
        "children": [{
            "FirstChild": "RECD19105-15-048",
            "children": [
                [{
                    ...
```



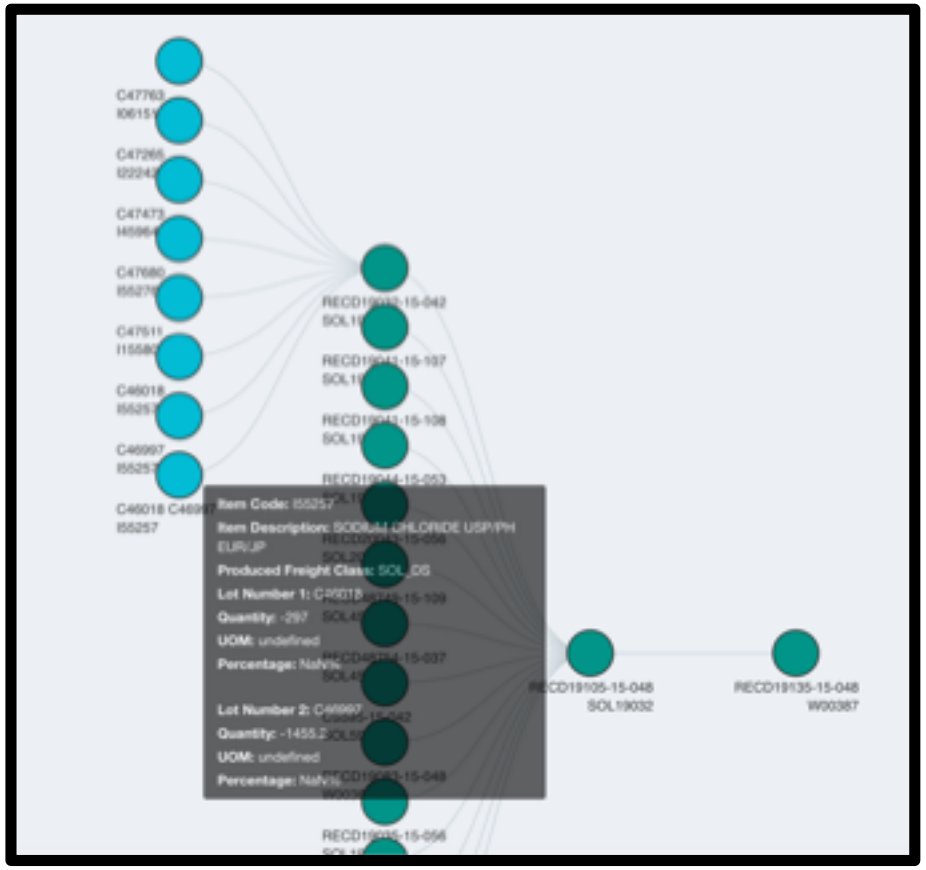
First Version – Core Functionality

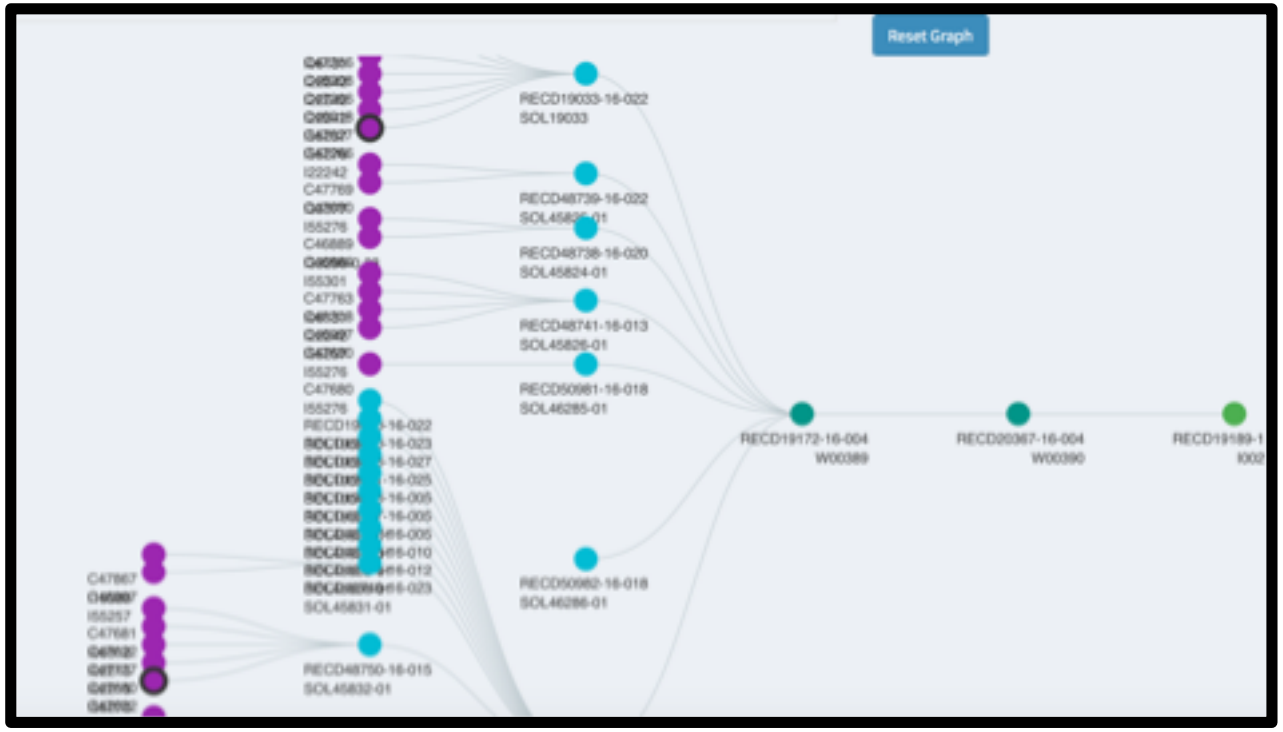


Technology Selection

- R: Open source statistical programming language, used to query data on raw materials, manufacturing pipelines, and transactions
- R Shiny: Interactive web application
- d3.js: Custom charting and visualizations that were embedded in R Shiny applications

Details Come Later





Shiny and d3 Communication

server.R: session$sendCustomMessage(type="jsondata",var_json)

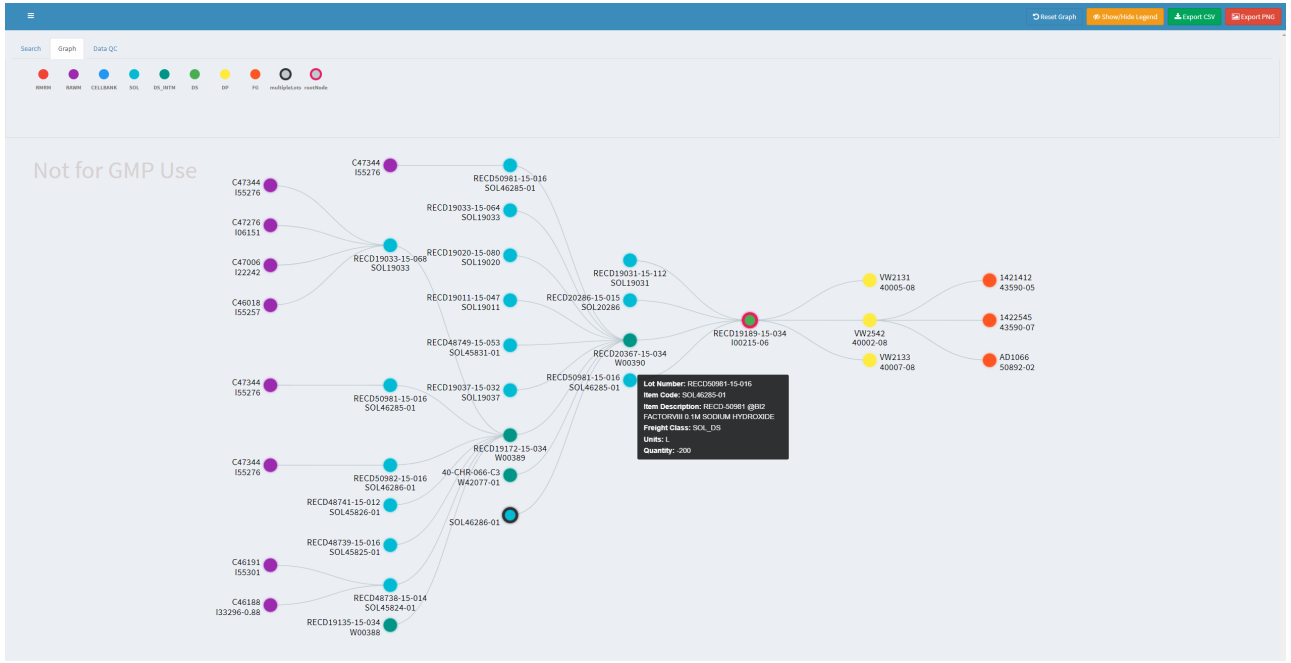
www/: main.js

```
 Shiny.addCustomMessageHandler("jsondata", function (message) {
        if (typeof(message) !== 'undefined') {
            var json_data = JSON.parse(message);
            initTree(json_data.left);
            initSide(json_data.right);
        }
    });
```

ui.R: tags$script(src=”main.js")

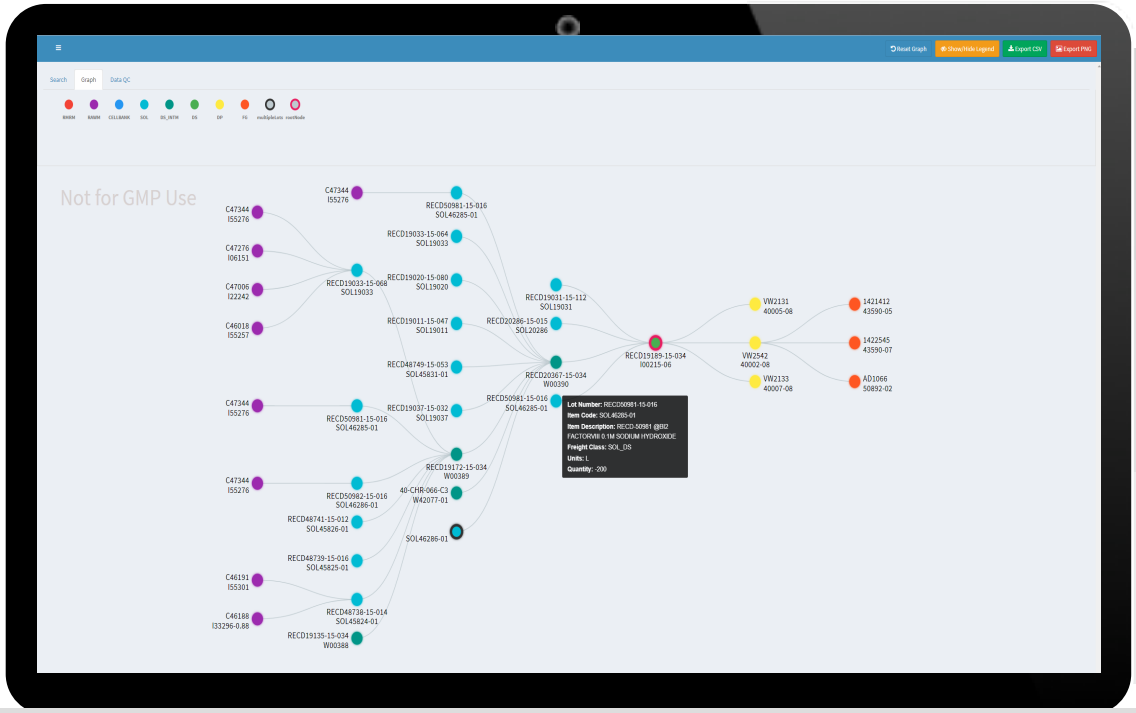
http://myinspirationinformation.com/visualisation/d3-js/integrating-d3-js-into-r-shiny/

Solution Overview



- Genealogy visualization and inspection solution for drug manufacturing quality control
- Centralized data mapping that links inputs to outputs, end-to-end across production
- Scientists can track the usage of any material throughout the product lifecycle
- Ability to quickly query for a specific drug recipe is critical to understanding trends and improving the at-line response time to potential negative deviations

Solution Overview

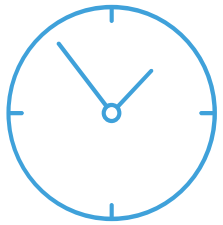


Intuitive node-based, drill-down interface, allows scientists and analysts to quickly identify a root lot associated with a deviation, the raw materials used in the solution, and any downstream substances or outputs that include the root lot.

To streamline the import of data from raw material vendors, data is now automatically stored in SQL Server, consolidated for analysis and visualization.

Genealogical data to may be quickly exported to build predictive models, to better understand the influence of specific raw materials on product quality attributes.

Results

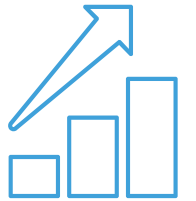


Time Savings

Whereas prior investigations could run on for months and required a large team of investigators, one user can now identify a problematic substance within a few days.

Investigator Expertise

Typically, mining this genealogy data is usually a task reserved for database users with advanced SQL skills. This solution eliminates the reliance on IT, enabling scientists with greater knowledge with the manufacturing pipeline to quickly investigate the raw material process.



Actionable Intelligence

Ultimately, this quality control application will be used to provide scientists with deeper insights into the manufacturing process, and the ability to track trends and report on large-scale production.