Selection and Implementation of a Global FTIR Database

Yann Bidault, Don Tucker, and Gregory M. Banik, Ph.D.,
Bio-Rad Laboratories, Inc., Informatics Division, Philadelphia, PA, USA

Spectroscopy 280019

Abstract

The KnowItAll cheminformatics software platform is a multi-technique environment that has been specifically designed to help spectroscopists and chemists solve a broad range of problems that they encounter in their daily work. Among many other tools, the system offers a chemical and spectral database visualizing, mining, and building module and an advanced database search module. Together, these tools help researchers to identify unknown compounds by comparing a query spectrum to matching spectra in "user-built" or commercial databases of reference spectra. KnowItAll's enterprise solutions (client-server) allow for deployment over multiple sites, so as to create a central resource that can be shared by users located in distinct and distant laboratories. Reference data access optimization, improved efficiency, and cost reduction are the main benefits that global companies can expect from such a global deployment.

This case study describes how a global chemical company utilized the KnowItAll platform to build a common Fourier transform infrared (FTIR) spectral resource and share it among sites spread over three continents: North America, Asia, and Europe. Challenges and measured performance will be presented, providing useful insights on organizational change requirements and benefits that can result from this approach.

Introduction

The company is a global manufacturer of engineered thermoplastic resins and material solutions for customers in a variety of industries. The company holds a strong position as a global engineering and development leader working to develop and validate new innovative concepts.

The spectroscopy team serves six chemistry-based polymer businesses, with nine technology sites worldwide including portions of two global research centers owned by a parent company. At time of trial and implementation, the spectroscopy team comprised six to seven Ph.D. analytical chemists and eight to ten spectroscopy technicians of diverse background. Typically, a laboratory receives a contaminated sample from a collection line, and technicians are asked to conclusively identify the contaminant. FTIR provides specific information about chemical bonding and molecular structures. It is widely used to determine the identity of unknown substances such as organic compounds and polymers and can also provide useful information on many inorganic materials. Identification of a contaminant can be made by comparing the contaminant spectrum to reference spectra, according to the following process: (1) contaminant isolation, (2) collection of the FTIR spectrum of the extract, (3) spectral similarity search against reference IR spectra collections.

Project Context and Analysis

Context Overview

At the time when the project was initiated, the company's spectral similarity search system offered three levels of consultation—local direct, local through an expert, and off-site through an expert—utilizing reference spectra either in digital or hardcopy format. The identified performance limitations of the system were local library holdings, technician experience, site age, and local data quality. This context is summarized in Table 1.

BIO-RAD

Process	Steps	Options
Data consultation	Initial approach	Local Library
	Other options	On-Site expert
		IR hard-copies
		Off-site Expert
Data returned		Good data
		Bad data
		Search abandoned
Limiting Factors		Local library holdings
		Technician experience
		Site age
		Local data quality

Table 1. Contaminant identification

Project Goal

The goal of the project was to build an internal, proprietary FTIR spectra database of proprietary raw materials, products, common contaminants, and additives that would be accessible to all global analytical technology researchers.

Challenge

Prior to implementing the project, several types of problems were known to affect the contaminant identification system. A six-sigma methodology was adopted to identify and correct them. The results of this analysis phase are presented in Table 2 to Table 4.

While some issues were unlikely to be impacted by the deployment of a global reference database, others were expected to be affected positively. Table 2 presents a list of the existing issues with an evaluation of the influence of the project implementation on their resolution.

Issue		Expected effect of deployment on issue
Wide variety of instruments in use		Neutral
Wide variety of skils/expertise areas among FT-IR users		Neutral
Hardcopy spectra often required for reference		Positive
Sporadic similarity searching practices among sites		Positive
Sites maintain own local libraries	Variable quantity and quality	Positive
	Include local raw materials/products only	Positive
~ 5% top search matches are incorrect	Multiple personnel involved in difficult interpretations	Neutral
	Increased sample turn around time - potential production loss	Positive
	Inaccurate/inconclusive results	Positive
File transfer between sites: requires additional manpower and opens communication gaps		Positive

Table 2. Process issues

Project Objectives

Seven project objectives to improve the contaminant identification process within the company were defined, as outlined in Table 3.

#	Objectives
1	Align spectral interpretation and identification efforts among sites
2	Eliminate dependence on hardcopy reference spectra where possible
3	Enable newer sites to efficiently utilize resources from more established laboratories
4	Alleviate sampling differences where possible
5	Provide single searching platform for various original formats
6	Facilitate knowledge transfer among locations and businesses
7	Allow room for future expansion

Table 3. Project objectives

Project Outline

Relying on the above analysis, a sequence of steps was created to outline the implementation process of the project (Table 4).

#	Step
1	Determine team members and configuration
2	Identify critical criteria - user friendliness, software, flexibility, etc.
3	Identify appropriate commercial packages
4	Determine IT limitations and requirements
5	Complete multi-site trial of chosen package(s)
6	Rank software based on critical criteria
7	Complete purchase of selected software
8	Global administrator trains administrators in library creation and searching
9	Administrators train users at each site
10	Global Administrator maintains library quality
11	Project success depends on search speed, database usage, and the amount/quality of uploaded data

Table 4. Project outline

Project Implementation

Team Demographics

Nine sites were involved in the project: five in the United States, two in Europe, one in China and one in India. The central server was deployed in at one of the US sites. The project team demographics were as follows:

Function	Characteristics
Global Administrator	Based in the US with server access
	Responsible for overall administration
	Liaison to software vendor
Administrators	One each location
	Identifies and uploads good quality spectra
	Trains users on site
	First point of contact for site-based questions
Users	Several each location
	Dependent upon site functions
	Includes spectroscopists, technicians, product developers, etc.
	General software use

Table 5. Team demographics

Solution Selection Criteria: Software Evaluations

Methodology

The choice of a software package to implement the project was based on team-determined criteria utilizing a simplified Quality Function Deployment (QFD) methodology. Four software packages were identified as possible solutions for the project. In a first step, a pre-trial evaluation was initiated in order to validate the QFD matrix and determine which packages would be further extensively tested.

Results

Table 6 and Figure 1 present the evaluation results. The performance level was defined as follows: H = High, M = Medium, L = Low. Based on pre-trial evaluation, software package 3 was eliminated. Software package 1, 2, and 4 displayed close matching pre-trial scores with regard to customer satisfaction criteria and were retained for subsequent trial.

Customer Expectation	PRE-TRIAL RESULTS						POST-TRIAL RESULTS					
	Importance	Package 1	Package 2	Package 3	Package 4	Total	Importance	Package 1	Package 2	Package 3	Package 4	Total
Possibility to create own database	5	H	H	L	H	140	9	H	H		H	243
Create various customer defined fields	5	H	H	L	H	140	9	H	M		H	189
Use multiple spectral databases, private & commercial	5	H	H	L	H	140	9	L	H		H	171
Search on spectrum (match spectrum)	5	H	H		H	135	9	H	H		H	243
Accurate matching	5	M	M		M	45	9	M	M		M	81
List of top matches	5	H	H		H	135	9	M	H		H	189
Variable matching mechanisms	5	H	H		M	105	9	M	M		H	135
Flexible searching: user defined fields, name,...	5	H	H	I	H	140	9	M	M		H	135
User friendly	5	M	M	M	M	60	9	L	M		H	117
Multiple file-format compatibility	5	H	H	m	H	150	9	H	H		H	243
Download spectra in numerical format	5	M	H	L	M	48	3	M	M		H	45
Fast searching	3	M	L	L	L	18	3	M	L		H	39
Simultaneous searching on multiple databases	3	M	H	L	H	66	3	M	H		H	63
Simultaneous searching on selected databases	3	M	H		H	63	3	M	H		H	63
Domain communication ability	1	L	L	L	L	4	1	L	L		H	11
Chemical Information system/tie-in	1	L				1	1	L				1
Cost of buying software	5	L	M	H	M	80	9	L	M		H	117
Cost of maintenance	5	L	L	L	L	20	9	L	L		L	27
Works via network	5	H	H	H	H	180	9	M	M		H	135
Expandable to other spectra (NMR,...)	0	H	M	M	H	0	9	M	L		H	117

Table 6. Software evaluation criteria and pre-trial evaluation

With regard to post-trial satisfaction results, package 1 and package 2 underperformed initial expectations. Package 4, however, over-performed for five criteria out of twenty.

Figure 1 is a graphical display of the results collected in Table 6, with pre-trial performance evaluation/expectation in blue, and post trial results in red. All tested packages display results that seem higher for post-trial than pre-trail testing. This is partly due to re-adjustment of the importance factor after the pre-trial phase. In reality, packages 1 and 2 globally underperformed compared to their pre-trial evaluation. Package 4, however, overperformed its pre-trail ranking and clearly outperformed the two other tested packages. Logically, software package 4—Bio-Rad's KnowItAll platform—was selected as the solution of choice to be deployed in the project.

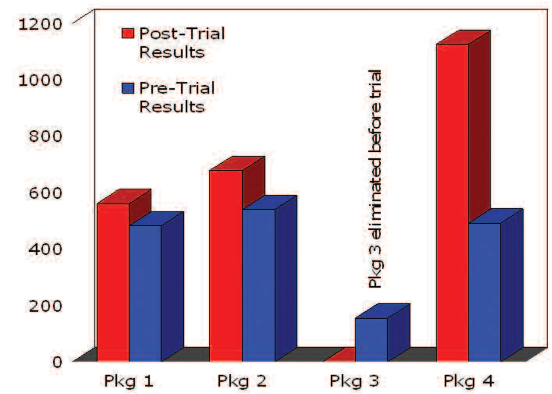

Fig. 1. Final comparison

Global Deployment

The KnowItAll solution was deployed on the eight sites mentioned above, the star indicating the location of the server.

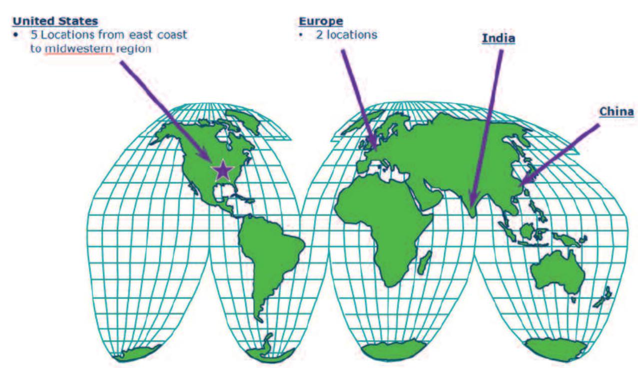

Fig. 2. Sites of KnowItAll deployment

When designing a central data resource, it becomes necessary to set some "global" or "multi-site" standardization rules with regard to criteria such as:

1. Software interface (identical or with some degree of personalization),
2. Spectrum (resolution, range, %Transmittance vs. Absorbance, ATR correction, etc.),
3. Database sub-grouping (per site, lab, sample type, technique),
4. Point in process, and
5. Naming conventions (for instance, "brown bag" vs. cellulose).

Results

Search Speed Evaluation

Search speed performance (project process step #11 in Table 4) was evaluated at each site. Each local administrator submitted 15 replicate search requests and recorded the times elapsed before receiving the results. Results are presented in Figure 3 and discussed below:

Summary: Search Speed Metrics		
Site	**Average (sec)**	**Std. Dev.**
India	2.07	0.26
US1	1.13	0.35
US2	2.16	0.73
China	2.77	0.91
Nether.	0.53	0.05
US3	2.33	0.49
US4	0.80	0.19

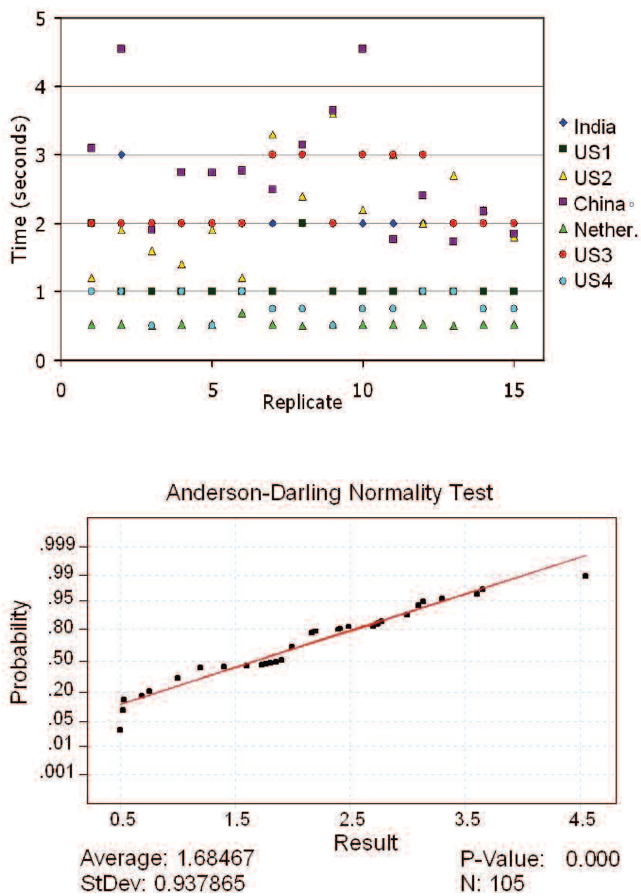

Fig. 3 — Search speed results

Europe site two and US site five were unable to perform testing due to local network issues at that time. Consequently, results are presented for seven sites out of nine. China and India were expected to experience longer search times due to the server location. Search speed performances exhibit search times that were widely scattered among the sites, but far below the 60 second upper specification limit.

Database Usage

Database usage was the second criterion of success listed in the Six-Sigma methodology. The following chart presents the number of server hits accessed through the network monthly for each site. One year after the project implementation (mid-2008), some changes in changes in the project and the personnel affectation decreased the number of overall users slightly – but the overall database usage remains at a high level.

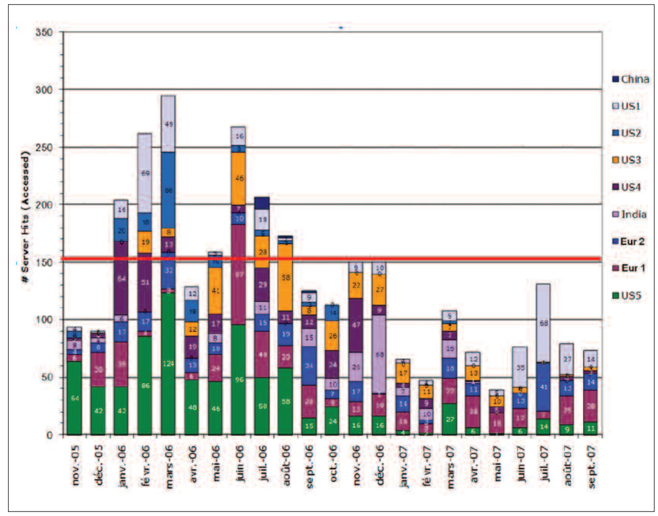

Fig. 4. — Database usage

Collection Building

The final success criterion was the amount of data uploaded in the shared collections as well as the quality of the newly entered data. Figure 5 displays a steady progression in the number of spectra entered and database collection during the project implementation. One year later, although the number of collections and spectra has not significantly changed, the quality of the data has improved.

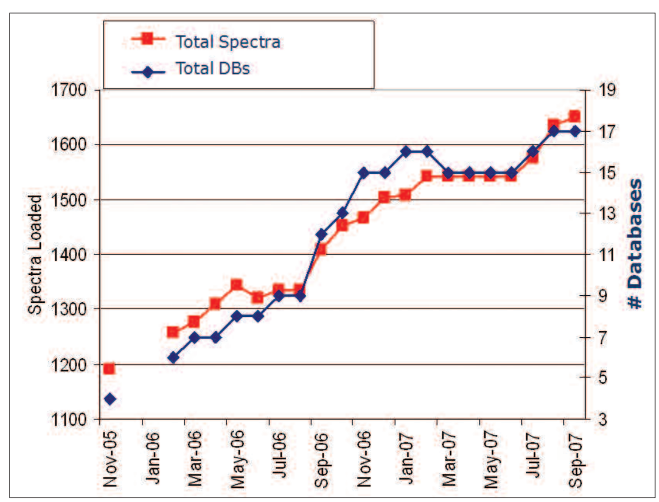

Fig. 5. — Database building

Project Success Evaluation and Conclusion

Some underestimated issues…

As in all new projects, some issues were unforeseen or underestimated. Some of them will need to be addressed by organizational changes:

- Database building can be time consuming, and is not usually measured by management; prioritization can be difficult.
- Administrator turnover is counterproductive.
- Technician buy-in is the true key to success. It is important to ensure that value is being added to their jobs.
- Navigating away from the instrument software in order to search other libraries is not ideal under normal circumstances.
- Software specific to a leased computer will have to be reinstalled on the new computer when the lease is over.
- High-quality spectra are critical for use, but adding commercial collections must be justified by high usage metrics

...but some business-critical problems solved!

Despite these issues, at least one business-critical issue has already been solved using the system. This increase in efficiency is a crucial outcome of this project. The expected savings are, in the company's own words, "staggering", and the potential applications to other areas are promising.

BIO-RAD

Bio-Rad Laboratories, Inc.

Informatics Division **Website:** www.knowitall.com **Contact Us:** www.knowitall.com/contactus

280019-REV20160505