Parallel bioreactor systems: integrated bioprocess information management

by Claudia M. Huether and Dr Falk Schneider

Bacteria and yeast, as well as animal and human cells, are widely used in research and industry. The cultivation of cells in bioreactors is a common method for growing and screening such cells, and for accelerating process development in biotechnology. Regulatory demands in regard to the manufacture, research and development of bioproducts, are steadily increasing. In particular, the automation of bioprocesses, as well as a comprehensive and standardised analysis and documentation are required. Advanced information management systems can meet these requirements. The following article will give an insight into the notable advantage of using process-integrated data management with parallel bioreactor systems.

The fermentation of microorganisms and eukaryotic cells in bioreactors is one important aspect of biotechnological research and production. Pharmaceutically relevant proteins, enzymes for biocatalytic use, as well as dietary supplements such as vitamins or antioxidants, are among the large range of products developed and manufactured in the pharma, chemistry and feed and food sectors using bioreactor technologies.

The complexity of biological systems, as well as increasing regulatory requirements (e.g. the FDA's Process Analytical Technology initiative - PAT) demand a more and more extensive, detailed and standardised acquisition and documentation of bioprocess data. Whereas ten years ago, fewer than fifteen different process data points were tracked during conventional bioreactor fermentations - now, as many as to forty data points are collected in a well equipped system. Moreover, descriptive data like the organism used for the process, the media composition, cultivation profiles or the achieved product yields need to be correlated and included into the analysis. Therefore, comprehensive and integrated tools for information management and evaluation are required.

Data generation in fermentation processes

During the complete runtime of a bioreactor cultivation process, data is generated. On the one hand, online process data such as temperature, pH level, dissolved oxygen, optical density, or exhaust gasses like oxygen and carbon dioxide, are continuously logged. Usually these parameters are recorded in short time intervals of minutes or even seconds. On the other hand, less frequently collected data from external analytical devices are recorded hourly or daily. This kind of data might describe cell numbers or concentrations of substrates, products or by-products. Data from external analysers is conventionally obtained in an offline manner by manual sampling and transferred to the control software as offline data. Open connectivity achieved by communication protocols like OPC (Openness, Productivity & Collaboration) find their way into pharmaceutical and biotech industries and tend to be the key to comprehensive process data management. By OPC integration of analysers, analytical data from samples can be automatically transferred to the system to be stored together with the primary process data. Combined with auto-sampling techniques, a fully automated sampling and even a close loop feed-back of analytical data is possible. Consequently, OPC technology opens the possibility for easy automation of a broad range of analyses and thereby complies with the requirements of the FDA's PAT initiative. Furthermore, any up-to-date information management approach has to consider the aspect of external data integration.

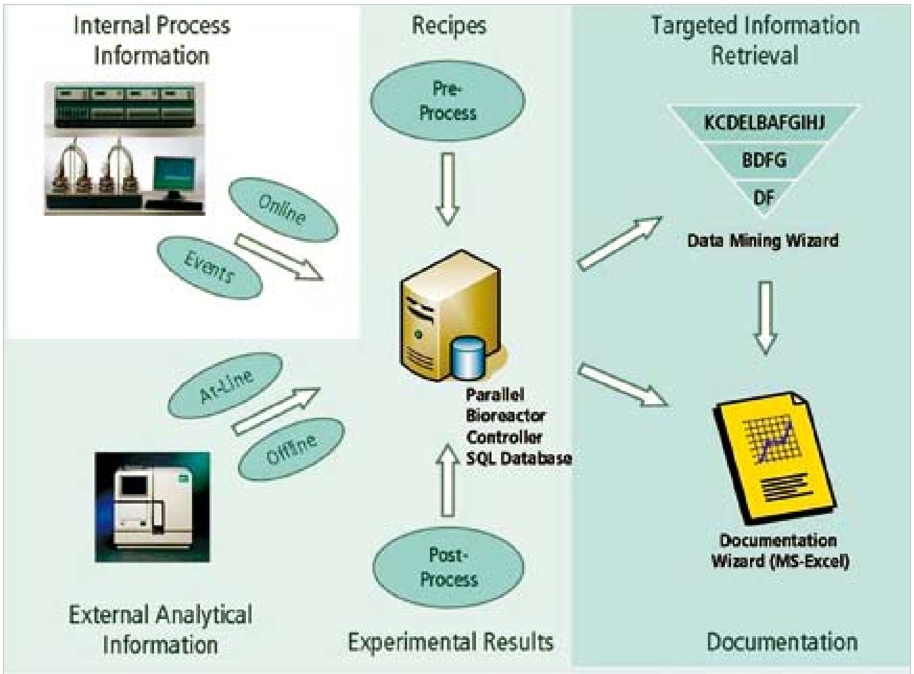
Systems running in parallel like the DASGIP Parallel Bioreactor System, where four to sixteen bioreactors are used at the same time, are widely accepted as powerful tools to provide a systematic approach in strain development, media or process optimisation. Industrial decision makers appreciate the time and laboratory savings, as well as the improved reproducibility. Furthermore, they benefit from direct comparability of process data between different bioreactors and from the opportunity to implement new technologies like Design of Experiment (DoE). However, these systems generate an enormous quantity of process data even in a single experimental cycle. For example, for a typical set of eight parallel microbial fermentations this means eight reactors with about 30 process data points per reactor to be recorded every 30 seconds. Thus, each of these tracks represents approximately 3000 values per day; for the sum of all reactors this is about 700,000 values. Based on an average fermentation runtime of two days per week, you get 1.4 million values per parallel run. This data amount equals roughly 10 MB of storage capacity. In cell culture applications the same amount of data is recieved: the longer mean runtime compared to microbial applications is compensated by less experiments per time and longer logging intervals. On average about 500 MB of raw process data is generated in the system's process database every year using such a parallel bioreactor system - even if data reduction techniques are used. In relation to today's hard-drive storage capacity the amount of data is no longer considered critical, however more and more attention has to be paid to the ability to retrieve information.

Focussing on a single laboratory, the bioreactor controller can be considered as the local data centre aimed to provide the necessary data mining functionality. On a more global scale, large process information software solutions with extended data mining capabilities, as well as historians with sophisticated data reduction algorithms, concentrate and harmonise data generated in process development, pilot scale and manufacturing. The definition of generalised tags for the data points in such a system is a challenge, as lots of widely spread workgroups have to be involved. Since the bioreactor controller is only a small data provider in such a system, OPC and open database connectivity (ODBC) are again the key for integration into these large process information solutions.

TURNING PROCESS DATA INTO VALUABLE PROCESS INFORMATION

Bioreactor control software traditionally only covers the recording of the raw process data, as described above, in combination with an event record. However, for later investigations of fermentation and cultivation processes, or to compare individual historical runs, it is neccessary to add more detailed process describing information into the analysis. First of all, general descriptions on the experimental setup e.g. the bacteria strain or cell line, the chemical composition of the cultivation media, nutrients utilised as feeds, controller set-points and parameters, as well as feeding profiles, have to be recorded during preparation [Figure 1]. Furthermore, the profound knowledge about post-process analytics e.g. achieved product yields, viable cell densities or a general quality valuation, turn raw process data into valuable process information and thereby allow comprehensive analysis.

This context information is usually not kept together with the runtime data but organised by the individual working groups in spreadsheets or sometimes in individual databases. Such information usually contains less than 100 values per



Timeline	Container	Type	Example	Data Points	Changes per Data Point	Information Coverage
Pre-Process	Recipe	Biological	Bacteria Strain, Cell Line	10	0	Process Information Management System
		Chemical	Media Composition	10	0	
		Process	Set Points, Feed Profiles, PID Parameter	10-100	1-5	
Process	Archive	Online	pH, DO, Temperature	30	100-10.000	Conventional Bioreactor Controler
		Event	Pressure Drop	n/a	n/a	
		Offline	OD_{600}, Offline pH, VCD	5	1-10	
		At-Line	Glucose Conc., Cell Numbers	5-50	10-100	
Post-Process	Lab-Notebook, Spreadsheet	Analytics	Product Yield	10	0	
		Quality	0% (useless)-100%	10	0	

Figure 1. Bioprocess information generated within a parallel cultivation process.

run and can be ignored - compared to the amount of process data generated. However, this particular information is the key to any targeted information retrieval or so called 'data mining'. Only the easy access to context information in combination with powerful query tools turns process data into valuable process information.

Information Management Tool for Parallel Bioreactor Systems

DASGIP has extended its Bioreactor Control Software by an integrated bioprocess information and documentation management tool, which particularly serves the demands of parallel bioreactor cultivations [Figure 1]. This tool ensures that all information during the whole lifecycle of a cultivation, beginning with the preparation (recipe management), continued by the process run and up to the post-process analytics, is collected and linked together. Online process data is completed with any user defined additional process information needed. A data mining wizard provides an intuitive query interface to select a set of runs that are automatically exported into a comprehensive MS-Excel worksheet for further analysis.

Recipe Manager

Prior to the fermentation or cultivation run, all static parameters of the experimental setup are defined as a recipe in an intuitive operable user interface. The recipe manager enables a smart compiling of individual parallel protocols clearly laid out in tabular form, as well as the convenient management of full sets of recipes where modifications like parameter variations are eased up by the use of a copy and paste functionality. During the run, the online process data is recorded and can be replenished with off-line data from external analytical devices (i.e. glucose concentration, OD, cell density). OPC connectivity opens the possibility for an at-line integration of external data, thus addressing different degrees of laboratory automation.

Data Mining Wizard

Upon completion of a bioreactor run, the collected process information can be rounded off by the addition of the post-process analytical data, such asthe product yield and maximum growth rate, etc. The integrated data mining wizard allows the target-oriented data retrieval. The complete database can be searched for all datasets sharing similar key parameters by applying a SQL (Structured Query Language) -like query interface. As shown in Figure 2, input fields in the column headers of the parameter table are used to enter the query criteria. To select, for example, all processes where the *E. coli* strain K12 was used for the production of a certain product achieving a product yield of at least 0.2 g/L, these two parameters are entered in the

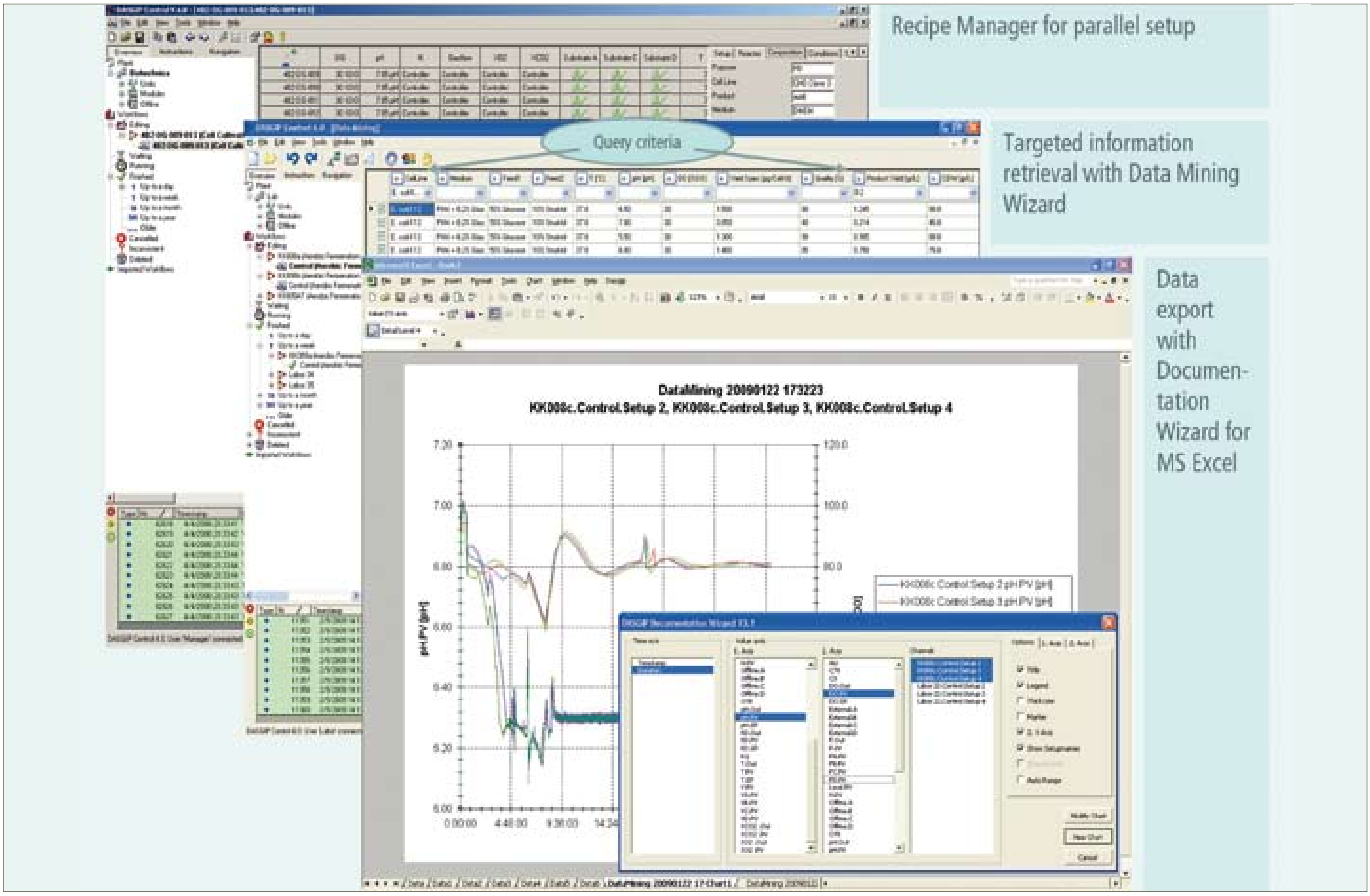


Figure 2. DASGIP Bioprocess Information Management System.

input fields for column 'Cell Line' and 'Yield', respectively. Step by step the table is reduced, showing only the runs matching the search criteria [Figure 2].

Documentation Wizard

All runs, selected as a query result, including their recipes, process data tracks, as well as events, and user defined content information, can be directly exported to MS-Excel by a single mouse click. The documentation wizard for MS-Excel generates a comprehensive documentation of each run. Online data tracks of current runs and historical runs can be compared to each other, whereas a chart creator automatically creates charts populated with the selected tracks (i.e. pH and dissolved oxygen) from a subset of runs.

The information management tool thereby ensures not only comparability of data generated in successive experiments or within one project but also makes it possible to analyse, at any time, historical data stored in the database under new foci.

Closing note

The increasing complexity of cultivation processes and new approaches like the FDA's PAT initiative has led to the need for comprehensive information management tools for the large amount of data generated by such biotechnological processes. Tools are needed that can store the online data along with descriptive key parameters from recipes, as well as external analytical data, derived either during or after the process. The integrated data management tool mentioned in this article provides a straight forward approach. On the one hand, it enables sophisticated software functionality for the control of small-scale parallel bioreactors. On the other hand, the OPC connectivity opens the integration into third-party supervisory process information systems.

The authors

Claudia M. Huether
Marketing and Communication
email: c.huether@dasgip.de

Dr Falk Schneider
Executive Vice-President &
Director Software Engineering
email: f.schneider@dasgip.de

DASGIP AG
Rudolf-Schulten-Str. 5
52428 Juelich,
Germany
web: www.dasgip.com

www.biotech-online.com & search 12181