Automation of Experimental Workflows for High Throughput Robotic Cultivations

Lucas Kaspersetz,a Fabian Schröder-Kleeberg,a Federico M. Mione,b Ernesto C. Martinez,b Peter Neubauer,a Mariano Nicolas Cruz-Bournazoua\*

Mariano.n.cruzbournazou@tu-berlin.de

Abstract

Robotic cultivation platforms optimally operated with computational tools can serve as generators of high-information content data. However, the generated data should be accompanied with the executed workflows to enable reproducibility of experiments. Workflow management systems support data provenance and enable traceability of experimental workflows. In contrast to production environments with highly standardized structures and strict operating regimes, integration of workflows in robotic experimental facilities is often challenging. The challenges are due to proprietary interfaces, rapidly changing objectives, and lack of adequate representation of the implemented workflows which prevents their interoperability and automated metadata capture. In this work, we present the design and implementation of a digital infrastructure with a workflow management system as the top layer. The execution of experimental workflows is automated using Apache Airflow enabling to manage all necessary steps for fed-batch cultivations, including sampling, sample transport by a mobile robot, feed additions, data collection, storage in a SQL database and model fitting. Thus, the generated experimental data can be accompanied with the executed workflows, facilitating transparency, and sharing of experimental protocols.

**Keywords**: directed acyclic graphs, workflow automation, reproducibility

* 1. Introduction

The lack of high quality, reproducible experimental data for biological systems is still the main obstacle to achieve (Rogers et al., 2022) scientific breakthroughs in bioengineering. To tackle this issue, robotic cultivation platforms optimally operated with computational tools can serve as generators of Findable, Accessible, Interoperable and Re-usable (FAIR) data with high information content. To this aim, the generated data should be accompanied by the executed workflows, which are necessary for increasing reuse and reproducibility (Mitchell et al., 2022). Particularly in bioprocess development, where scale-up decisions are based on small-scale experimental data, transparency and interoperability in experimental and computational workflows are essential and currently missing. A workflow management system (WMS) for automatically scheduling and executing the experimental and computational tasks not only increases the degree of automation but also contributes to implementing FAIR data principles. The implementation of a WMS is often challenging due to missing standardized device interfaces or lack of adequate data management infrastructure for handling the different data sources (Maffettone et al., 2023). Current approaches for the management and scheduling of workflows in laboratories are often limited to subtasks such as analytical workflows, using mobile robots for automated sample transport (Neubert et al., 2019) or specific scheduling tasks of parallel cultivations (Bromig et al., 2022). In such cases, not all tasks, constraints, or data processing steps that are necessary for the effective operation of robotic cultivations can be automated or digitally handled for reproducibility and provenance. The open-source platform Apache Airflow (AA) provides the possibility of developing, scheduling, and monitoring workflows via directed acyclic graphs (DAG) (Harenslak and Ruiter, 2021). The applicability of DAGs for enforcing FAIR principles by design in high throughput cultivation facilities has been recently addressed for computational workflows (Mione et al., 2022). Still, without a similar approach for the experimental tasks, FAIRness cannot be achieved. In this work, we present the design and implementation of a WMS based on DAGs in a robotic cultivation facility. We show that the WMS not only increases the degree of automation but also contributes to FAIRizing experimental workflows. We demonstrate the feasibility and added value of our tool with *E. coli* BL21 (DE3) scale-down experiments for the production of elastin like proteins (Huber et al., 2014).

* 1. Material and Methods
		1. BioXplorer cultivation platform

The BioXplorer 100 (H.E.L group, London, United Kingdom) composed of eight parallel cultivations in glass STRs, equipped with off-gas analyzers (BlueVary, BlueSens, Herten, Germany) was used for the cultivation workflows. The system is integrated into a liquid handling station (Tecan Group, Männedorf, Switzerland). A mobile robotic lab assistant (Astechproject Ltd., Runcorn, United Kingdom) was used for automated sample transport. The reader is referred to (Kaspersetz et al., 2022) for a detailed description of the platform and sampling procedures and at-line measurement procedures.

* + 1. Strain

All experiments were carried out with *E. coli* BL21 (DE3) strains, carrying the pET28-NMBL-mEGFP-TEVrec-(V2Y)15-His expressing a recombinant fusion protein of an elastin like protein and eGFP, under the IPTG inducible *lac*UV5‐promoter.

* + 1. Cultivation

The preculture was set to an OD600 of 0.25 and cultured in 50 mL EnPresso B medium (Enpresso GmbH, Berlin, Germany) with 6 U L−1 Reagent A in 500 mL shake flask overnight at 37°C and 220 rpm in an orbital shaker (25 mm amplitude, Adolf Kühner AG, Birsfelden, Switzerland). Main-cultures were run in four parallel glass stirred tank reactors (STR) each equipped with one Rushton type impeller at 37°C and pH was controlled at 7.0 with 7.5 % (NH3(aq)). The main cultures were started as 90 mL batch cultures with an initial glucose concentration of 5 g L-1. Two substrate addition strategies were used in the scale-down experiment: continuous feeding and bolus feeding profiles. The pulse-based feed followed a 10~min interval. Aeration and stirring were increased following a pre-defined scheme.

* + 1. Apache Airflow

Apache Airflow 2.2.4 and Python 3.7 were used to programmatically describe, schedule and monitor the experimental cultivation workflows. The official Docker Image for Apache Airflow is hosted on DockerHub (apache/airflow:2.2.4). The workflows were authored as DAGs. All necessary for running an experimental workflow were included in docker-compose.yml. All containerized applications were run via Docker Desktop 4.4.4 (73704) for Windows and a Windows Subsystem Linux 2 based back engine. The code is publicly available under: <https://git.tu-berlin.de/bvt-htbd/airflow_workflow/-/tree/escape?ref_type=heads>.

* + 1. E. coli growth model

The model is based on a mechanistic *E. coli* model with glucose partitioning, overflow metabolism, and acetate re-cycling. The mathematical model was formulated as an ODE system describing the changes in state variables for glucose, acetate, biomass, product and dissolved oxygen tension (DOT). The parameters of the model were obtained by fitting the model to the experimental data. A more detailed description of the underlying model and the functioning of the framework can be found in (Kim et al., 2023).

* 1. Device integration and workflow automation

In bioprocess development, the implementation of automated cultivation workflows requires the integration of different devices, including mobile robots. Such robotic cultivation platforms usually consist of parallel bioreactor systems embedded in liquid handling stations (LHS) and analytical devices such as high-throughput analysers or additional LHS. In Fig. 1 a hierarchical infrastructure for the robotic cultivation platform is presented. All devices were integrated following a client-server architecture based on google Remote Procedure Call (gRPC) or Standardization in Lab Automation (SiLA2). The WMS in AA was implemented as the top layer. This digital infrastructure aims to provide a comprehensive tracking of the complete experimental workflow and to support transparency and reproducibility (Mitchell et al., 2022). All process data were stored in a SQL-database, while a shared file system was used to exchange data between containerized applications. The Apache Airflow web-service served as a user-interface for the experimental operator, allowing to trigger and monitor all experimental workflow steps.



Figure 1: Schematic representation of the robotic cultivation platform with a workflow management system as the top layer.

* + 1. Implementation of experimental workflows

All experimental workflows for managing fed-batch cultivations were represented through DAGs. The Airflow scheduler executed the respective tasks on an array of workers (fig. 1) while following the pre-defined dependencies. The automated sampling DAG manages the sampling event of the LHS, the transportation of the microtiter plate by the mobile robot and the corresponding analytics. All nodes which are associated to a device, used the PythonOperator and the corresponding client for communication. The DAG starts with a node that triggers the sampling procedure of the LHS. If the sampling event is successfully completed, the transportation node is carried out. After completion of the sample transport the corresponding analytics device is started, while in parallel the LHS can perform an additional task. Once the at-line analytics has been completed, the database is accessed to retrieve the corresponding results. The logging and monitoring DAG manages the data needed by the cultivation system as well as queries data from the database for monitoring and updating the parameter estimation procedures. For the cultivation system and the off-gas analysis the implementation followed the same approach, using the PythonOperator and a client-server architecture. As soon as both devices are requested to send their online measurements to the database, the TriggerDagRunOperator starts the parameter estimation DAG. The computational pipeline for parameter estimation is executed inside a docker container. Containerized applications reduce integration effort, whilst increasing interoperability and reproducibility of computational workflows (Boettiger, 2015).



Figure 2: Representation of robotic cultivation workflows as directed acyclic graphs.

* 1. Case study – Scale-down cultivation

In order to demonstrate the practicality of the system, the proposed infrastructure was used to manage the experimental workflows for conducting scale-down fed-batch cultivations. Process performance parameters such as yields and titres of biotechnological process are often faltered when scaled-up to industrial scale. This limitation occurs due to lack of robustness of the microbial host to perturbations in large-scale conditions (Olsson et al., 2022). As cells move through the industrial-scale reactor, they are steadily exposed to changing environments. Hence, investigating the microbial response to such perturbations in usually homogenous and well-mixed small-scale reactors is necessary for a robust bioprocess development. Two different feeding regimes were applied, following a continuous feeding or a pulse-based feeding regime to investigate the robustness of *E. coli* BL21 (DE3) to glucose oscillations. In Fig. 3 the cultivation data for parallel fed-batch cultivations with continuous feeding (R1 & R2) and pulse-based feeding (R3 & R4) are shown. The WMS successfully collected and stored all online process measurements in the database. The sampling DAG allowed for seamless acquisition of at-line measurements, while controlling the mobile robot for automated sample transport. The pulse-based feeding was scheduled over a 10-min cycle interval which leads to an oscillating pattern in the measured O2-concentration. For the assessed feeding strategies, no influence on the max. specific substrate uptake rate (*qS*max = 1.5g g h-1) and specific product (fig. 3) parameters were observed. However, an increased maintenance coefficient (*q*m = 0.09 g g h-1) for the pulse-based feeding strategy was observed.



Figure 3: Parallel *E. coli* fed-batch cultivations, producing elastin like proteins, with continuous feeding (R1 & R2) in comparison to pulse-based feeding (R3 & R4).

* 1. Conclusions

We present the successful implementation of a WMS, based on DAGs, in a robotic cultivation facility. The proposed digital infrastructure is capable of scheduling and managing all necessary workflows for scale-down fed-batch cultivations. In the scale-down experiment, *E. coli* BL21 showed to be robust with no adverse measurable physiological responses to glucose oscillations under the tested conditions. The generated experimental data can be accompanied with the executed DAGs, which facilitates sharing of the conducted workflows, while increasing reproducibility and FAIRness of complex high-throughput experiments.

**Acknowledgements**

We gratefully acknowledge the financial support of the German Federal Ministry of Education and Research (01DD20002A – KIWI biolab).

References

Boettiger, C., 2015. An introduction to Docker for reproducible research. ACM SIGOPS Oper. Syst. Rev. 49, 71–79.

Bromig, L., von den Eichen, N., Weuster-Botz, D., 2022. Control of parallelized bioreactors I: dynamic scheduling software for efficient bioprocess management in high-throughput systems. Bioprocess Biosyst. Eng. 45, 1927–1937.

Harenslak, B., Ruiter, J. de, 2021. Data pipelines with Apache Airflow. Manning Publications Co, Shelter Island, NY.

Huber, M.C., Schreiber, A., Wild, W., Benz, K., Schiller, S.M., 2014. Introducing a combinatorial DNA-toolbox platform constituting defined protein-based biohybrid-materials. Biomaterials 35, 8767–8779.

Kaspersetz, L., Waldburger, S., Schermeyer, M.-T., Riedel, S.L., Groß, S., Neubauer, P., Cruz-Bournazou, M.-N., 2022. Automated Bioprocess Feedback Operation in a High-Throughput Facility via the Integration of a Mobile Robotic Lab Assistant. Front. Chem. Eng. 4.

Kim, J.W., Krausch, N., Aizpuru, J., Barz, T., Lucia, S., Neubauer, P., Cruz Bournazou, M.N., 2023. Model predictive control and moving horizon estimation for adaptive optimal bolus feeding in high-throughput cultivation of E. coli. Comput. Chem. Eng. 172, 108158.

Mione, F.M., Silva, A.N., Luna, M.F., Cruz B., M.N., Martinez, E.C., 2022. Managing Experimental-Computational Workflows in Robotic Platforms using Directed Acyclic Graphs. In: Yamashita, Y., Kano, M. (Eds.), Computer Aided Chemical Engineering, 14 International Symposium on Process Systems Engineering. Elsevier, pp. 1495–1500.

Mitchell, S.N., Lahiff, A., Cummings, N., Hollocombe, J., Boskamp, B., Field, R., Reddyhoff, D., Zarebski, K., Wilson, A., Viola, B., Burke, M., Archibald, B., Bessell, P., Blackwell, R., Boden, L.A., Brett, A., Brett, S., Dundas, R., Enright, J., Gonzalez-Beltran, A.N., Harris, C., Hinder, I., David Hughes, C., Knight, M., Mano, V., McMonagle, C., Mellor, D., Mohr, S., Marion, G., Matthews, L., McKendrick, I.J., Mark Pooley, C., Porphyre, T., Reeves, A., Townsend, E., Turner, R., Walton, J., Reeve, R., 2022. FAIR data pipeline: provenance-driven data management for traceable scientific workflows. Philos. Trans. R. Soc. Math. Phys. Eng. Sci. 380, 20210300.

Neubert, S., Gu, X., Göde, B., Roddelkopf, T., Fleischer, H., Stoll, N., Thurow, K., 2019. Workflow Management System for the Integration of Mobile Robots in Future Labs of Life Sciences. Chem. Ing. Tech. 91, 294–304.

Rogers, A.W., Vega‐Ramon, F., Yan, J., Río‐Chanona, E.A., Jing, K., Zhang, D., 2022. A transfer learning approach for predictive modeling of bioprocesses using small data. Biotechnol. Bioeng. 119, 411–422.