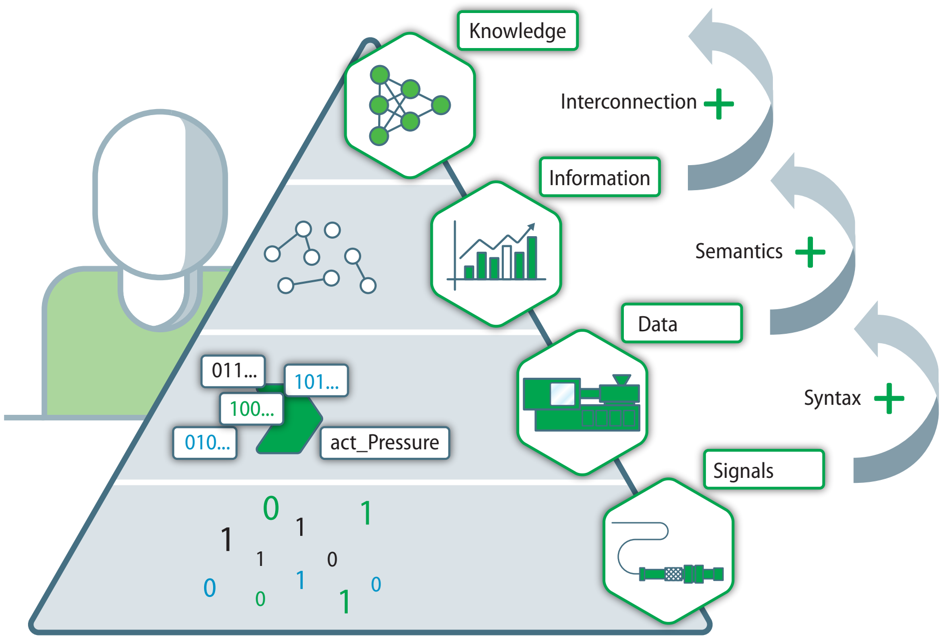
Part 9: Consistent Data Structures for Sustainable Access to Secure Knowledge

Complete Data without Ifs and Buts

In production and research, process data and previously defined quality values are tracked to determine optimization models and pursue consistent quality management. In contrast, data that is not relevant to the specific application is rarely recorded. To ensure long-term access to complete and contextualized trial data, research data management in conjunction with process data acquisition is required and initialized at the IKV as part of the Plastics Innovation Center 4.0.



Structure and interconnection along the information-pyramid subsequently generate knowledge from sensor signals. Source: IKV; graphic: © Hanser

Statistical design and analysis of experiments are central tools in the engineering sciences. In research as well as in industrial setup-procedures, the influence of specific machine-settings on process and quality variables can be investigated with little experimental and measurement effort. Based on research questions and hypotheses, monitoring variables are defined, such as characteristic values for injection pressure or the machine-side screw-acceleration-curves. Real or simulated trials then evaluate the hypothesis. Although valid results are generated concerning the parameter space that allow a statement to be made about the initial hypothesis, the data collected also allow a large number of other correlations and dependencies to be analyzed that are usually not the focus of concern in the actual case. The research infrastructure of the Plastics Innovation Center 4.0 (PIC 4.0) now shows how the consistency of process-data can be ensured with the aid of structured data-acquisition and digital services in trial and production planning.

Between Hypothesis and Gained Knowledge

The importance of consistent data and reliable data-formats usually raises after a process has been completed. On the one hand, data such as the sensor-signal of a cavity-pressure-sensor, the screw-position of an injection unit or the simulated course of a flow-front-velocity are available, which must first be synchronized and correlated [1]. Different cycle-counters or timestamps may be available in the respective systems. On the other hand, there is also master data on the machine, the mold and the material, a current configuration of the system-technology as well as notes recorded with paper and pen, which are relevant for the significance or the reconstruction of the test series.

If this information gets lost completely or just partially, the results can only be reused to a limited extent outside the specific research-question of concern. Trials with the same mold and the same material even on the same machine must be repeated in case of doubt because, for example, the screw movement was not relevant in previous investigations and was therefore not recorded or there is uncertainty about individual setting values. Data acquisition in PIC 4.0 prevents this by defining a standard dataset that is stored in parallel with focused trials.

Aimless Data-Acquisition, Unused Data-Graveyards

Consistent data are also important for machine learning methods. Incomplete datasets or those that cannot be unambiguously assigned often limit the direct applicability in industrial production or result in inaccuracy of the derived models [2]. Aimless data-acquisition, on the

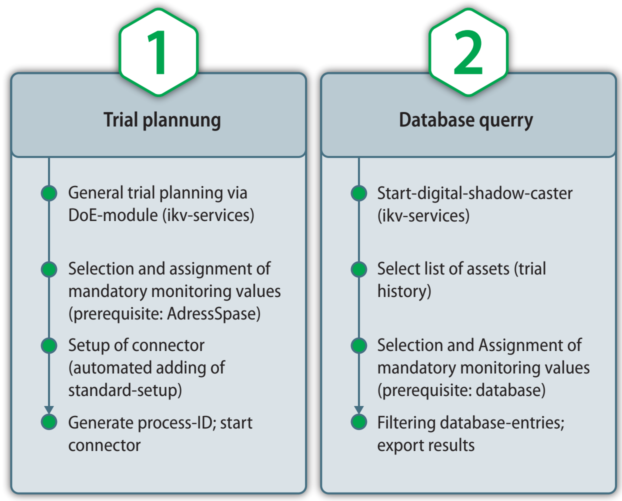


Fig. 1. The two scenarios of targeted data assignment: perspective from experimental design (left) or retrospective via database query (right). Source: IKV; graphic: © Hanser

other hand, leads to unused data graveyards. Setup-procedures open up a good analogy to the experimental trial plans used in research.

Valuable insights into how the setting variables on the production machines depend on each other enable model-based setup-support for consistent quality-assurance. However, the initial challenge is to set up data-acquisition in a targeted manner and to make data available for statistical or data-driven models from the field of artificial intelligence for model-validation long after the original trials. For this purpose, completeness, and the unambiguous assignment of quality variables to setting variables must be ensured.

Increased Efficiency through Standardized Test Planning

Research-data-management covers a broad range of topics. It ranges from the efficient data-acquisition of heterogeneous systems to the fast and simple processing of data into information to the long-term access of generated knowledge while simultaneously linking the relevant data (**Title figure**) [3]. For the data and information management of a research infrastructure such as that of PIC 4.0, this means specifically keeping track of data from trials consistently and being able to access it beyond the initial research project, for example for model validation in other projects.

The challenge is to keep track of all the process-settings required for reliable reconstruction of the trial data and to assign all the trials to specific trial plans and projects. Two scenarios must be distinguished in the functional sequence for this purpose (**Fig. 1**). On the one hand, a proactive definition must be made based on the project-specific trial planning, and on the other hand, a retrospective database-query must be made from historical data.

A service platform based on open-source-technologies (ikv-services) developed within the project already supports users in generating statistical experimental designs (Design of Experiments, DoE) in a standardized way [4]. Another module allows the data-acquisition-client (Connector) to be configured by selecting the monitoring variables desired for the particular series of trials (**Fig. 2**). The Connector queries the monitored quantities and sends the numerical values to an internal broker-network, where they are, among others, stored in a database.

In addition to the individually specifiable target-variables, the set of monitoring variables is expanded by a basic selection that includes all relevant process-settings and helpful measurement-diagrams for screw-movement, even if these are not the focus of the investigations in the specific project. »

Digital Shadow

The digital shadow is the core-concept of the Cluster of Excellence "Internet of Production". It describes a semantic aggregation of relevant data traces from different sources to enable a data-based mapping of an object's (asset's) behavior in production.

Info

Text

Univ.-Prof. Dr.-Ing. Christian Hopmann has held the Chair of Plastics Processing as Head of the Institute for Plastics Processing (IVK) in Industry and Craft at RWTH Aachen University, Germany, since 2011.
Pascal Bibow, M.Sc. RWTH, has been working since September 2016 on intelligent production systems at IKV. Since January 2020, he has been Managing Director of the Plastics Innovation Center 4.0; pascal.bibow@ikv.rwth-aachen.de
Dipl.-Ing. Mauritius Schmitz has been Scientific Director for Digitization and Industry 4.0 since January 2019 at the IKV and has been responsible coordinator of the institute's research activities in the cluster of excellence Internet of Production.

Acknowledgments

The construction of the Plastics Innovation Center 4.0 is funded by the State of North Rhine-Westphalia and by the European Regional Development Fund (ERDF).

The Series Continues

Das PIC 4.0 is close to completion and the implementation of machine and plant technology is scheduled. In the last article of the series, read about the installed possibilities offered by the PIC 4.0 and its research fields in the future. It appears in one of the next issues.

References & Digital Version

You can find the list of references and a PDF file of the article at ***www.kunststoffe-international.com/archive***

German Version

Read the German version of the article in our magazine ***Kunststoffe*** or at ***www.kunststoffe.de***

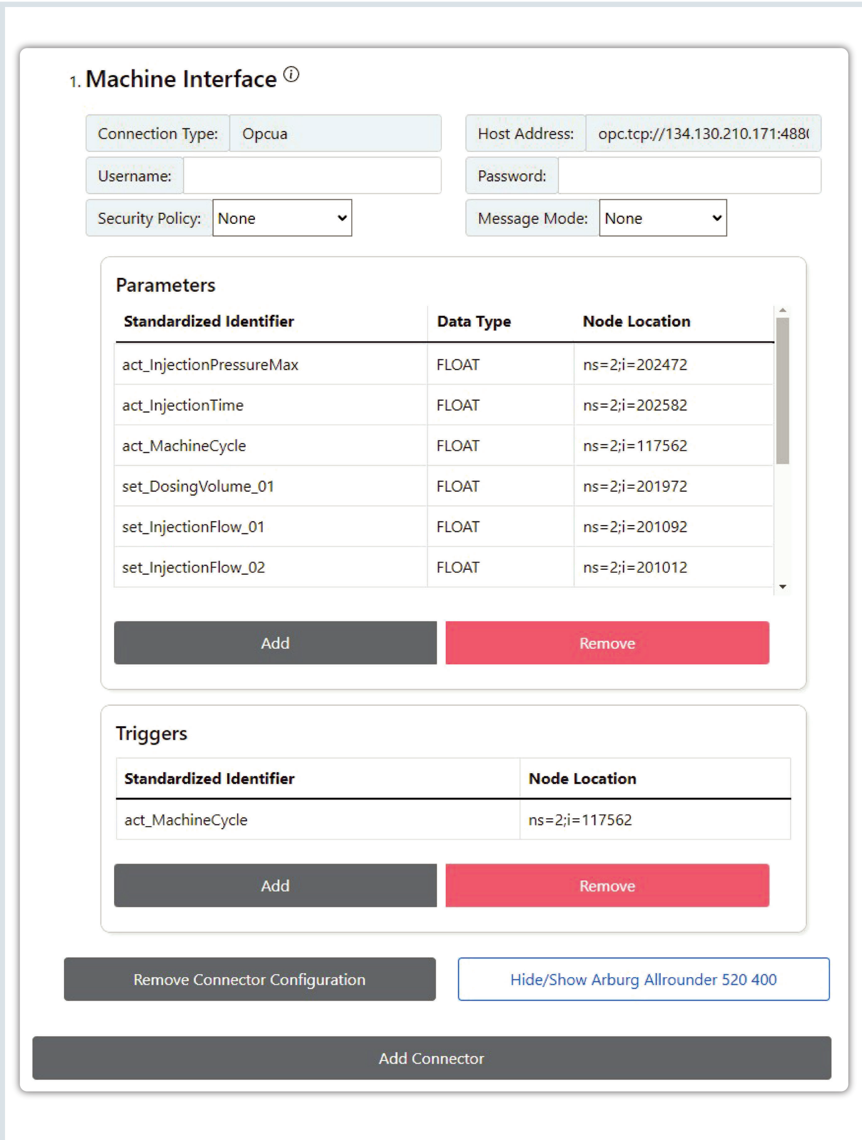


Fig. 2. In the user interface for manual Connector configuration (excerpt), project-specific monitoring variables can be individually compiled. Source: IKV; Graphic: © Hanser

Finally, when the Connector is started, a process-ID is generated that allows data-traces from different devices to be retrospectively assigned to a project. Analog and unstructured data-packages, for example handwritten test-notes or error-logs, can also be attributed to the trials retrospectively.

Data Traces of the Process History

The second application-scenario concerns the database-query on processes carried out in the past. For this purpose, the database can be searched for the measured variables; however, it is more purposeful to compile the data-traces to the setting-combinations searched for. The concept of the digital shadow developed in the "Internet of Production" cluster of excellence can be implemented and demonstrated close to the application for contextualizing the data within PIC 4.0 [5]. In the form of a further service for the available platform, a functionality is currently being developed that determines and outputs the data-traces of the process-history based on setting-combinations of the process to be defined (digital shadow caster).

With reference to the metadata generated for the processes, as well as the process-ID, corresponding devices and plants can initially be selected as data sources (assets). From these, a user is to define the data sources to be included. Subsequently, the available parameters are to be displayed; the basic configuration is ensured by the trial-planning-scenario. Subsequently, a relevant subset of the data-sets suitable for the intended analysis or model-validation can be exported via an automated database-filter-command.

Deep Dive into the Datalake

The need for intelligent filtering techniques arises from the various databases and database-structures that are operated when setting up data-acquisition and for the semantic correlation of individual data-traces. While cycle-related numerical values or character strings can be written into a relational database, metadata (for example master data) on the machine and further equipment used or on the process-configuration are stored in non-relational databases. Since each production system generates its own data traces as a data source, these can each be stored in separate tables in relational databases, while the associated master data can be stored in collections in the non-relational database. The correlation of data traces and metadata can then only take place via unique identifiers such as an order number in an industrial production or the process-ID introduced here.

While the basic collection, storage and archiving of even large volumes of data is no longer a technical problem, it is often still a long way to go before the data can be processed in a targeted manner and any monetary benefit can be generated. The scenarios presented for setting up project-related data-acquisition and retrospective correlation of historical data are intended to generate long-term added value. In addition to the increased transparency and retrievability, this is also expressed in a guaranteed consistency and completeness of the data – for this purpose, further software-services are being developed in addition to the existing ones. In this way, data will also be made usable for answering further questions, such as the transferability of research approaches to different machines and equipment or the processing of other materials. ■