Supplementary Information (SI) for Digital Discovery.
This journal is © The Royal Society of Chemistry 2026

Supplementary Information

WeChemSyn: Semantic Modeling of Wet-Chemical Syntheses in a Self-Driving Lab for Nano- and Advanced Materials

Markus Schilling,[a] Harald Bresch,[b] Bernd Bayerlein,*[a] and Bastian Ruehle*[b]

a. Federal Institute for Materials Research and Testing (BAM), Unter den Eichen 87, 12205 Berlin, Germany
b. Federal Institute for Materials Research and Testing (BAM), Richard-Willstaetter-Str. 11, 12489 Berlin, Germany.
* Correspondance: bernd.bayerlein@bam.de, bastian.ruehle@bam.de

---

In practice, constructing application- and domain-level ontologies is often accomplished by reusing, in a targeted manner, concepts from established ontologies rather than defining all terms from scratch. Leveraging these mature resources enables consistent representation of experimental methods, investigation metadata, process-chemistry descriptions, qualities, units, and laboratory-centric entities. Table 1 compiles selected ontologies that can be reused and briefly explains how each may support the modeling of wet-chemical nanomaterial synthesis workflows in SDL settings.

Table S1. Selected reusable ontologies and their relevance for wet chemical nanomaterial synthesis in a self-driving laboratory.

Ontology	***Primary scope***	***Relevance***
CHMO[1] (Chemical Methods Ontology)	Methods that cover chemical experimentation, which include data collection methods (e.g., microscopy, spectroscopy), sample preparation/separation, and synthesis methods, plus related instruments.	Provides a controlled vocabulary for characterization and synthesis methods that appear in automated workflows and in instrument metadata that SDL pipelines produce.
CAO[2] (Chemical Analysis Ontology)	Vocabulary for chemical analysis metadata that is organized under a BFO framework, with emphasis on concepts, material entities, information content entities, roles, and processes.	Supports structured descriptions of analysis contexts and data items that accompany characterization outputs, which helps standardize how analytical results and metadata get annotated.
PATO[3] (Phenotype and Trait Ontology)	Ontology of qualities/attributes (properties) that supports consistent modeling of qualities that inhere in entities.	Provides reusable quality terms (e.g., size, morphology-related qualities) that can support consistent descriptions of nanomaterial properties and qualitative outcomes across experiments.

[1] https://bioportal.bioontology.org/ontologies/CHMO (accessed 01/28/2026)
[2] https://fgiovannetti.github.io/cao/ (accessed 01/28/2026)
[3] https://obofoundry.org/ontology/pato.html (accessed 01/28/2026)

Ontology	*Primary scope*	*Relevance*
PROCO[4] (Process Chemistry Ontology)	Formal ontology for process chemistry that targets the development and optimization of production processes and scale-up considerations.	Contributes terminology that supports synthesis workflow descriptions and process-centric metadata, which complements SDL goals that target process optimization and reproducibility.
AFO[5] (Allotrope Foundation Ontologies)	Ontology suite that provides a vocabulary and semantic model for laboratory analytical processes and that aligns to BFO, with core domains that include equipment, material, process, and results.	Offers lab-informatics–oriented concepts that map naturally to automated laboratory execution, instrument outputs, and structured results, which supports interoperability with lab data standards and tooling.
NCIt[6] (NCI Thesaurus / NCIT)	Broad biomedical reference terminology that supports clinical care and translational/basic research, with rich concept metadata and cross-links.	Provides widely used identifiers and synonyms for chemicals, biological concepts, and biomedical context that may appear in nano-bio interfaces, safety contexts, or application-driven experimental designs.
QUDT[7] (Quantities, Units, Dimensions and Types)	Unified representation of quantities, units, dimensions, and data types that supports explicit unit semantics and conversion metadata.	Enables explicit, machine-actionable representations of measurement values and units, which reduces ambiguity in SDL data streams and supports consistent aggregation across instruments.
OBI[8] (Ontology for Biomedical Investigations)	Integrated ontology for investigations, which covers study designs, protocols, instrumentation, materials, data generation, and analysis.	Provides investigation-centric modeling patterns (plans, assays, devices, data outputs) that support reproducible descriptions of experimental workflows, which map well to SDL "design–execute–report" cycles.

Example workflow: From natural language input to a populated and annotated Knowledge Graph representation

In the following, an example is given how an (imprecise and incomplete) sentence from a synthesis procedure can be exported as a knowledge graph that follows the WCSO using a simple, graphical tool (node editor), without requiring any deep knowledge of the underlying ontological concepts.

For the sentence

"The solution was heated at 80 C for 12 hours, while reagentX was added dropwise."

[4] https://bioportal.bioontology.org/ontologies/PROCO (accessed 01/28/2026)
[5] https://www.allotrope.org/ontologies (accessed 01/28/2026)
[6] https://bioportal.bioontology.org/ontologies/NCIT (accessed 01/28/2026)
[7] https://qudt.org/ (accessed 01/28/2026)
[8] https://obi-ontology.org/ (accessed 01/28/2026)

the LED-Base-16384_Llama[9] model generates the following intermediate action graph:

<HEAT> 80 °C <WAIT> 12 h <ADD> reagentX dropwise

which then gets directly further parsed by the node editor for creating the corresponding node graph, without requiring any additional user input. While parsing the action graph, the node editor outputs the following warning messages:

WARNING: Imprecise Addition Rate description found: dropwise. Assuming 1 mL/min.
WARNING: No stirring speed specified for STIR step. Using default stirring speed: 300 rpm
INFORMATION: Encountered an ADD step after a HEAT step. Automatically merging into INFUSE_WHILE_HEATING step.

The node graph that the node editor created from this sentence consists of a "Container Node", a "Chemical Node" for "reagentX", and an "Infuse While Heating Node" that uses the values of "80 °C" and "12 h" that were explicitly specified, as well as "300 rpm" from the built-in heuristics for the missing stirring speed and "1 mL/min" from the built-in heuristics for the addition rate (see red arrows in Figure S1).

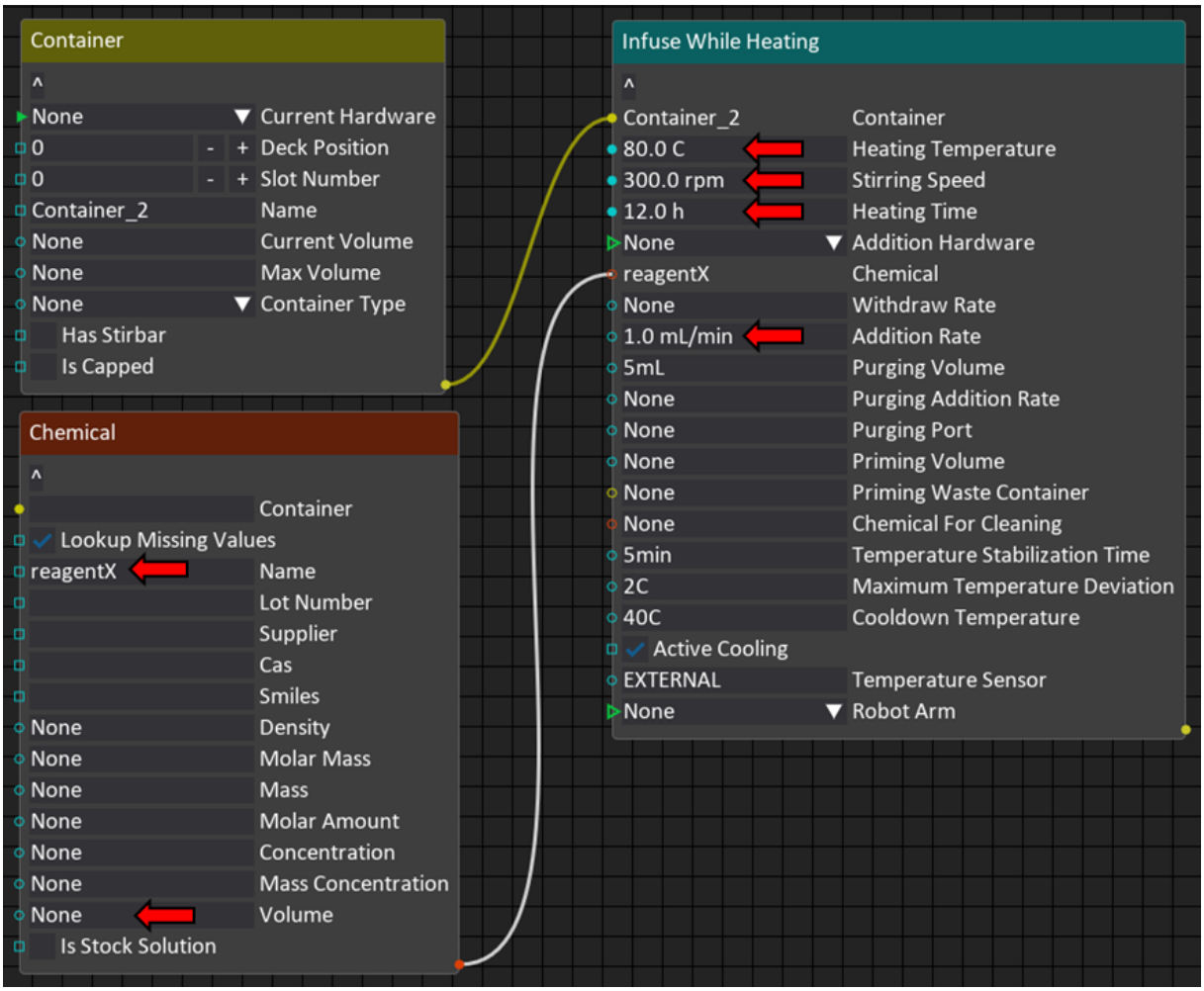

Figure S1. Automatically generated node graph for the sentence "*The solution was heated at 80 C for 12 hours, while reagentX was added dropwise.*", using the LED-Base-16384_Llama model[9]. The red arrows indicate the values that were automatically filled in, based on information from the text or using heuristics. For the missing volume, no value is generated (the field remains set to "None"), this needs to be rectified by the user before running the reaction.

Note that this is only one possible interpretation of the ambiguous procedure described by this sentence. It is not entirely clear whether the addition was evenly spaced out over the whole 12 hours, whether the solution was added

[9] https://huggingface.co/bruehle/LED-Base-16384_Llama (3/23/26); https://doi.org/10.5281/zenodo.15228014

dropwise, and then continued to stir for 12 hours after the addition was completed, or whether the 12 hours are included in the time it took to add reagentX dropwise. Also, no amount is given for reagentX. If the values chosen from heuristics do not represent the user intent, they can however still be easily corrected manually in the node graph. When exporting this node setup as a knowledge graph, the correct concepts from the T-Box (creating "SVS-patterns", annotating the heating and adding steps as concurrent) are applied and the A-Box entries are created automatically (see Table S2.). Only exemplary entries from the .ttl file are shown in the table, e.g., only one SVS pattern for the addition rate (similar patterns are created for all values, e.g., Temperature, Speed, Time) and also only one combination for the simultaneous processes during the infuse_while_heating step are shown (the .ttl file annotates all possible combinations as simultaneous)

Table S2. Excerpt from the automatically generated .ttl file when exporting the node graph shown in Figure S1 (gray background). Bold font face was added for better readability. Lines in italic with white background are human-readable substitutions of the IRIs used in the ttl file.

Automatically generated process attributes and SVS patterns:

<https://w3id.org/wcso/wcso_e/**infuse_while_heating**_511_UUID-41a11342-0866-4e01-a507-8f7a8afe39d9>	<https://w3id.org/pmd/co/**PMD_0000009**>	<https://w3id.org/wcso/wcso_e/infuse_while_heating_511_**addition_rate_process_attribute**_UUID-838e473d-4538-4a37-9729-561cc970a669>
infuse_while_heating	*has_process_attribute*	*addition_rate_process_attribute*
<https://w3id.org/wcso/wcso_e/infuse_while_heating_511_**addition_rate_process_attribute**_UUID-838e473d-4538-4a37-9729-561cc970a669>	<http://www.w3.org/1999/02/22-rdf-syntax-ns#**type**>	<https://w3id.org/pmd/co/**PMD_0000008**>
addition_rate_process_attribute	*is_a*	*process_attribute*
<https://w3id.org/wcso/wcso_e/infuse_while_heating_511_**addition_rate_process_attribute**_UUID-838e473d-4538-4a37-9729-561cc970a669>	<https://w3id.org/pmd/co/**PMD_0020127**>	<https://w3id.org/wcso/wcso_e/infuse_while_heating_511_**addition_rate**_UUID-655206ea-0822-4712-a6c0-23ecda0e45d7>
addition_rate_process_attribute	*refers_to*	*addition_rate*
<https://w3id.org/wcso/wcso_e/infuse_while_heating_511_**addition_rate**_UUID-655206ea-0822-4712-a6c0-23ecda0e45d7>	<http://www.w3.org/1999/02/22-rdf-syntax-ns#**type**>	<https://w3id.org/wcso/**nano_0000035**>
addition_rate	*is_a*	*addition_rate*
<https://w3id.org/wcso/wcso_e/infuse_while_heating_511_**addition_rate**_UUID-655206ea-0822-4712-a6c0-23ecda0e45d7>	<http://purl.obolibrary.org/obo/**IAO_0000419**>	<https://w3id.org/wcso/wcso_e/infuse_while_heating_511_**addition_rate_svs**_UUID-5b9ea33f-9c45-4ebb-9404-170f94b0fbfe>
addition_rate	*quality_is_specified_as*	*addition_rate_svs*
<https://w3id.org/wcso/wcso_e/infuse_while_heating_511_**addition_rate_svs**_UUID-5b9ea33f-9c45-4ebb-9404-170f94b0fbfe>	<http://www.w3.org/1999/02/22-rdf-syntax-ns#**type**>	<http://purl.obolibrary.org/obo/**OBI_0001931**>
addition_rate_svs	*is_a*	*scalar_value_specification*
<https://w3id.org/wcso/wcso_e/infuse_while_heating_511_**addition_rate_svs**_UUID-5b9ea33f-9c45-4ebb-9404-170f94b0fbfe>	<https://w3id.org/pmd/co/**PMD_0000020**>	<https://qudt.org/vocab/unit/**MilliL-PER-MIN**>
addition_rate_svs	*has_measurement_unit_label*	*MilliLiter-per-Minute*
<https://w3id.org/wcso/wcso_e/infuse_while_heating_511_**addition_rate_svs**_UUID-5b9ea33f-9c45-4ebb-9404-170f94b0fbfe>	<https://w3id.org/pmd/co/**PMD_0000006**>	**1.0**^^<http://www.w3.org/2001/XMLSchema#float>
addition_rate_svs	*has_value*	*1.0*

Automatically annotating an infuse_while_heating step as a "compound step" which has three temporal parts that are all simultaneous

<https://w3id.org/wcso/wcso_e/**infuse_while_heating**_511_UUID-41a11342-0866-4e01-a507-8f7a8afe39d9>	<http://purl.obolibrary.org/obo/**BFO_0000121**>	<https://w3id.org/wcso/wcso_e/infuse_while_heating_511_**heatingpart**_UUID-e3842f1b-a851-4dea-8f7b-0a27a29585ef>
infuse_while_heating	*has_temporal_part*	*heatingpart*
<https://w3id.org/wcso/wcso_e/**infuse_while_heating**_511_UUID-41a11342-0866-4e01-a507-8f7a8afe39d9>	<http://purl.obolibrary.org/obo/**BFO_0000121**>	<https://w3id.org/wcso/wcso_e/infuse_while_heating_511_**stirringpart**_UUID-75332dc7-c37c-4f89-8b9b-4e16f137b916>
infuse_while_heating	*has_temporal_part*	*stirringpart*
<https://w3id.org/wcso/wcso_e/**infuse_while_heating**_511_UUID-41a11342-0866-4e01-a507-8f7a8afe39d9>	<http://purl.obolibrary.org/obo/**BFO_0000121**>	<https://w3id.org/wcso/wcso_e/infuse_while_heating_511_**addingpart**_UUID-63b58cff-6ddb-4900-8d5c-6aac729b2e86>
infuse_while_heating	*has_temporal_part*	*addingpart*
<https://w3id.org/wcso/wcso_e/infuse_while_heating_511_**heatingpart**_UUID-e3842f1b-a851-4dea-8f7b-0a27a29585ef>	http://purl.obolibrary.org/obo/**RO_0002082**	<https://w3id.org/wcso/wcso_e/infuse_while_heating_511_**stirringpart**_UUID-75332dc7-c37c-4f89-8b9b-4e16f137b916>
heatingpart	*simultaneous_with*	*stirringpart*
<https://w3id.org/wcso/wcso_e/infuse_while_heating_511_**heatingpart**_UUID-e3842f1b-a851-4dea-8f7b-0a27a29585ef>	http://purl.obolibrary.org/obo/**RO_0002082**	<https://w3id.org/wcso/wcso_e/infuse_while_heating_511_**addingpart**_UUID-63b58cff-6ddb-4900-8d5c-6aac729b2e86>
heatingpart	*simultaneous_with*	*addingpart*

Note there are also several other fields in the created nodes in Figure S1, which either have their default values or are not populated at all. Many of these will be populated at run time, when the orchestrating backend applies further heuristics. These include for example which robot arm to use for putting the sample on the hotplate, which addition hardware, withdraw rate, purging rate, and purging port to use (depending on the volume of the syringe that is installed and the valve that is used), and also CAS and SMILES identifiers for the chemical (if it can be found on PubChem). The knowledge graph that can be constructed after a reaction was actually run on the SDL platform will therefore include much more information and enable others to unambiguously describe and reproduce the reaction in their labs.