

1 **Abstraction hierarchy to define biofoundry workflows and operations for**
2 **interoperable synthetic biology research and applications**

3
4 Haseong Kim^{1,2,3,4}, Nathan J. Hillson^{5,6}, Byung-Kwan Cho^{4,7}, Bong Hyun Sung^{1,2,3,4}, Dae-Hee
5 Lee^{1,2,3,4,8}, Dong-Myung Kim⁹, Min-Kyu Oh¹⁰, Matthew Wook Chang^{11,12,13,14}, Yong-Su Jin^{15,16},
6 Susan J Rosser^{17,18}, Peter Vegh^{17,18}, Rennos Fragkoudis^{17,18}, Rosalind Le Feuvre¹⁹, Nigel S.
7 Scrutton¹⁹, Marko Storch^{20,21}, Wonjae Seong^{1,2}, Paul S Freemont^{*20,21,22}, Seung-Goo Lee^{*1,2,3,4}

8
9 ¹Korea Biofoundry, Korea Research Institute of Bioscience and Biotechnology (KRIBB), Daejeon 34141, Republic
10 of Korea

11 ²Synthetic Biology Research Center, Korea Research Institute of Bioscience and Biotechnology (KRIBB), Daejeon
12 34141, Republic of Korea

13 ³Department of Biosystems and Bioengineering, KRIBB School of Biotechnology, University of Science and
14 Technology (UST), Daejeon 34113, Republic of Korea

15 ⁴Graduate School of Engineering Biology, Korea Advanced Institute of Science and Technology (KAIST), Daejeon
16 34141, Republic of Korea

17 ⁵DOE Agile BioFoundry, Emeryville, CA 94608 U.S.A.

18 ⁶Biological Systems and Engineering Division, Lawrence Berkeley National Lab, Berkeley, CA 94720 U.S.A

19 ⁷Department of Biological Sciences, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 34141,
20 Republic of Korea

21 ⁸Department of Integrative Biotechnology, College of Biotechnology and Bioengineering, Sungkyunkwan
22 University, Suwon, Gyeonggi-do 16419, Republic of Korea

23 ⁹Department of Chemical Engineering and Applied Chemistry, Chungnam National University, 99 Daehak-ro,
24 Daejeon, 34134, Republic of Korea

25 ¹⁰Department of Chemical and Biological Engineering, Korea University, Seongbuk-gu, Seoul 02841, Republic of
26 Korea

27 ¹¹NUS Synthetic Biology for Clinical and Technological Innovation (SynCTI), National University of Singapore,
28 Singapore

29 ¹²Synthetic Biology Translational Research Programme, Yong Loo Lin School of Medicine, National University of
30 Singapore, Singapore

31 ¹³Department of Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore, Singapore

32 ¹⁴National Centre for Engineering Biology (NCEB), Singapore, Singapore

33 ¹⁵Department of Food Science and Human Nutrition, University of Illinois at Urbana-Champaign, Urbana, IL 61801,
34 U.S.A.

35 ¹⁶Carl R. Woese Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, U.S.A.

36 ¹⁷School of Biological Sciences, University of Edinburgh, Edinburgh, United Kingdom

37 ¹⁸Edinburgh Genome Foundry, Centre for Engineering Biology, University of Edinburgh, Edinburgh, United
38 Kingdom

39 ¹⁹Manchester Institute of Biotechnology, University of Manchester (UoM), United Kingdom

40 ²⁰Department of Infectious Disease, Imperial College London, London, United Kingdom

41 ²¹London Biofoundry, Innovation Hub, White City Campus, Imperial College London, United Kingdom

42 ²²Care Research, and Technology Centre, UK Dementia Research Institute, London, United Kingdom

43

44 **Abstract**

45 Lack of standardization in biofoundries limits the scalability and efficiency of synthetic biology
46 research. Here, we propose an abstraction hierarchy that organizes biofoundry activities into four
47 interoperable levels: Project, Service/Capability, Workflow, and Unit Operation, effectively
48 streamlining the Design-Build-Test-Learn (DBTL) cycle. This framework enables more modular,
49 flexible, and automated experimental workflows. It improves communication between researchers
50 and systems, supports reproducibility, and facilitates better integration of software tools and
51 artificial intelligence. Our approach lays the foundation for a globally interoperable biofoundry
52 network, advancing collaborative synthetic biology and accelerating innovation in response to
53 scientific and societal challenges.

54

55 **Introduction and Motivations**

56 In June 2018, fifteen non-commercial biofoundries from four continents gathered in London and
57 agreed to establish the Global Biofoundry Alliance (GBA)¹, a collaborative effort to share
58 experiences and resources while addressing common challenges and unmet scientific and
59 engineering needs. Following the experience of the pandemic², the importance of biofoundries as
60 a main workforce of biomanufacturing and a sustainable bioeconomy has become even more
61 highlighted. Biofoundries are more than facilities for conducting experiments using automated
62 equipment; they are structured Research and Development (R&D) systems where biological
63 design, validated construction, functional assessment, and mathematical modeling are performed
64 following the Design-Build-Test-Learn (DBTL) engineering cycle¹. A biofoundry can be used for
65 conducting many heterologous experiments, necessitating the analysis of a wide range of different
66 experimental protocols and biological assays. In synthetic biology and engineering biology various
67 terms may be used interchangeably (and occasionally inappropriately), such as “protocols”,
68 “Standard Operating Procedures (SOPs)”, “workflows”, and “tasks”. Or, for example, the term
69 “protein design” sometimes refers only to the design step but at other times it can refer to the entire
70 DBTL process of protein design and engineering. For the operation of automated systems like
71 biofoundries, it is essential to precisely define these concepts and scope of terms used to describe
72 different biofoundry activities. Synthetic biology is an applied field that merges disciplines from

73 the life sciences and engineering, including molecular biology, chemical biology, genetics,
74 bioinformatics, chemical and computer engineering. The experiments conducted in biofoundries
75 extend beyond normal molecular and cell biology experiments and encompass a wide range of
76 application-driven protocols and methods. This diversity and complexity underscore the need for
77 a unified framework that not only standardized terminologies and methodologies but also
78 facilitates the exchange of best practices across biofoundries³. Therefore, it is timely to build an
79 international collaborative network for sharing biofoundry methodologies and applications using
80 common terminology and standardized methods.

81
82 Given that biofoundry workflows span from low-throughput manual protocols to high-throughput
83 operations using 96-, 384-, and 1536-well plates, quantitative metrics are crucial for benchmarking
84 performance improvements, ensuring reproducibility, and maintaining operational quality across
85 scales. These metrics also enable performance comparisons across different biofoundries, whether
86 the processes involve semi-automated workflows with manual plate transfers between instruments
87 or fully automated workflows using robotic arms⁴. However, developing such quantitative metrics
88 requires a foundational framework based on standardized protocols. Once standardized workflows
89 are established, biofoundries can create reference materials and calibration tools to assess
90 reproducibility and quality levels, enabling measurement comparisons across different instruments.
91 Prioritizing the standardization of workflows as a prerequisite for metric development enhances
92 the reliability and interoperability of biofoundry operations. This approach not only ensures
93 consistent performance across facilities but also mitigates the adverse effects of monopolies by
94 equipment manufacturers, fostering a more collaborative and equitable biofoundry ecosystem.

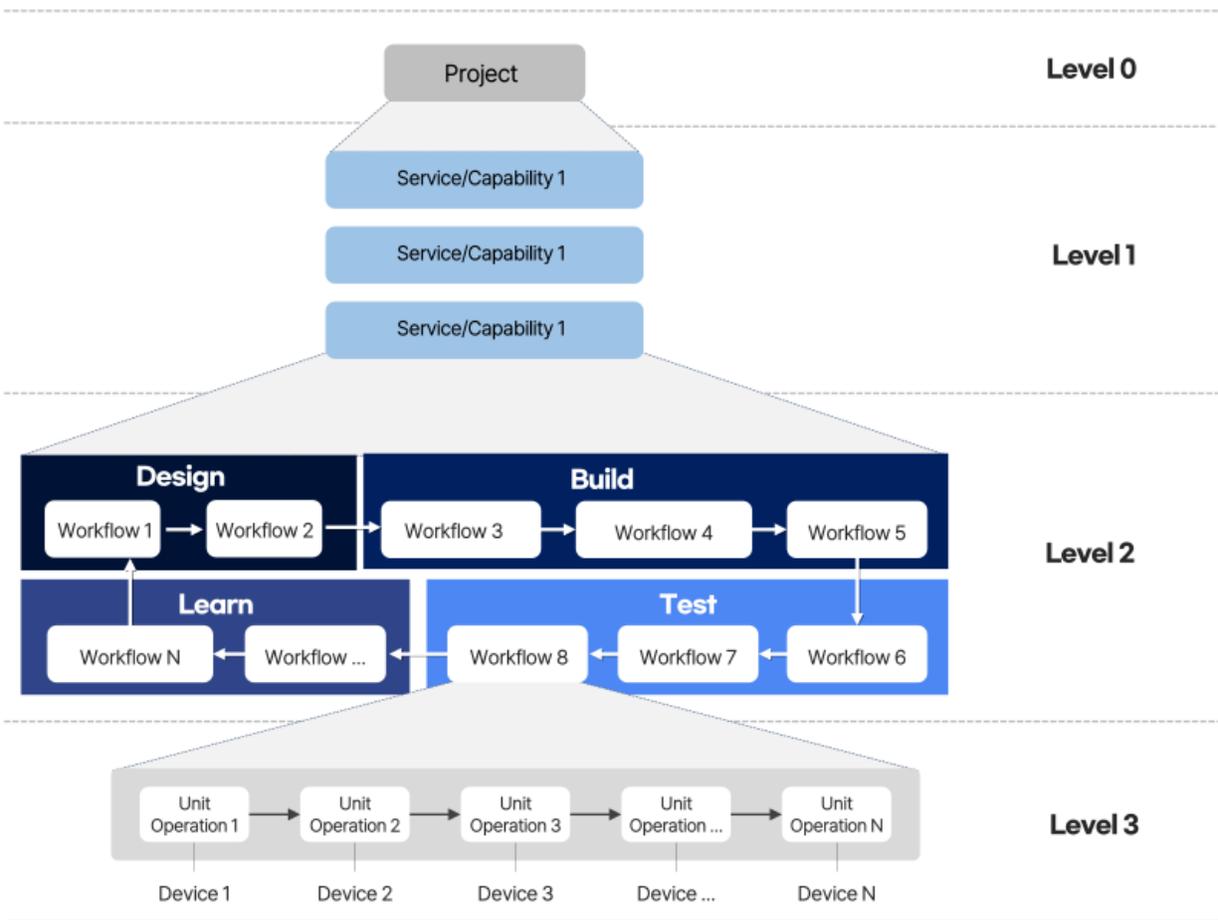
95
96 Shifting to a biofoundry environment introduces challenges in adapting experimental protocols.
97 Many existing lab-based synthetic biology protocols are optimized for manual execution and often
98 omit details that are assumed to be obvious to trained researchers. When these protocols are
99 directly applied to automated biofoundry platforms, which typically operate in 96/384-well plate
100 formats and use liquid-handling robots, differences in sample volumes, concentrations, and
101 equipment specifications can result in deviations from expected outcomes. In other words,
102 protocols that work reliably in manual settings may yield inconsistent results in automated
103 environments unless they are explicitly adapted for such systems. Additionally, human-executed
104 protocols often omit obvious steps in publications or laboratory manuals, such as sample
105 preparation. Automated workflows, however, require precise definitions of the location, state,
106 quantity, and behavior of all materials used. The same equipment is used differently depending on
107 the application, and equipment turnover in which older instruments are replaced by new ones,
108 further complicates reproducibility. These challenges underscore the need for highly abstracted
109 workflows that encapsulate biofoundry-specific processes while accommodating automation
110 variability.

111

112 **Abstraction Hierarchy for Biofoundry Operations**

113 To address the issues of biofoundry interoperability, we have designed a flexible abstraction
 114 hierarchy for the operation of a biofoundry (Figure 1). Level 0 refers to the Project that is to be
 115 carried out in the biofoundry. This represents a series of tasks to fulfill the requirements of external
 116 users who wish to use the biofoundry. Level 1 Service/Capability, refers to the functions that
 117 external users require from the biofoundry and/or that the biofoundry can provide. Level 2,
 118 Workflow, refers to the DBTL-based sequence of tasks needed to deliver the Service/Capability.
 119 Each workflow is intentionally assigned to a single stage of the DBTL cycle to ensure modularity
 120 and clarity in execution. Level 3 is Unit-operations which represents the actual hardware or
 121 software that will perform the tasks required to fulfill the desired workflow. Engineers or biologists
 122 working at the highest abstraction level do not need to understand the lowest Level 3 operations.

123



124

125 Figure 1. Abstraction hierarchy of biofoundry operations across four levels: Project (Level 0),
 126 Service/Capability (Level 1), Workflow (Level 2), and Unit Operation (Level 3). Each workflow
 127 corresponds to a modular step in the DBTL cycle and consists of linked unit operations mapped to
 128 devices. The diagram highlights how project goals are translated into executable protocols,
 129 ensuring clarity and interoperability from high-level intent to low-level execution.

130

131 *Level 1: Services and Capabilities*

132 Researchers and companies in the field of biotechnology can leverage the specialized services and
 133 capabilities provided by biofoundries to achieve their R&D project goals. Examples include
 134 modular long-DNA assembly or Artificial Intelligence (AI) driven protein engineering. In this
 135 report, a biofoundry capability refers to the specialized processes or activities conducted by
 136 biofoundries where clients can be from both academia and industry (including startups/spinouts,
 137 SMEs and larger organizations). Biofoundry services can be divided into various tiers - these range
 138 from simply providing access to specialist equipment to offering a fully comprehensive support
 139 package from project conception to commercialization and scale-up. We can categorize these tiers
 140 of services/capabilities in relation to the synthetic biology DBTL cycle (Table 1).

141

142 Table 1. Biofoundry service/capability category and examples

Type	Description	Examples
Tier 1	A service that supports the use of individual piece(s) of automated equipment.	Access to liquid handling robots for training users.
Tier 2	A service focusing on an individual stage of the DBTL cycle.	Though most biofoundry services require the combination of two or more stages in the DBTL cycle, Tier 2 is focused on activities related to a single stage. For example, a biofoundry provides a protein sequence library designed by ProteinMPNN ⁵ .
Tier 3	A service combining two or more DBTL stages such as DB, BT, TL, or LD.	Most of the heavily used services in the biofoundry belong to this tier. For example, AI model (L) training followed by protein design (D). If target gene sequence and structure are provided; the service of “protein library construction” involves simple construction (B) and sequence verification (T).
Tier 4	A service supporting the full DBTL cycle.	Example projects could include applying the full DBTL cycle to conduct research projects such as “Greenhouse gas bioconversion enzyme discovery and engineering”; “Plastic degradation microorganism engineering”; “Production of functional materials for food/medicine” etc. A good example of the DBTL cycle in Tier 4 is demonstrated by the SYNBIOCHEM Biofoundry ⁶ , which highlights the power of biofoundries in discovering novel chemical pathways and optimizing product titer during early-stage scale-up. In the healthcare sector, high-demand areas such as Cell Line Development and Antibody Discovery could also serve as Tier 4 examples.

143

144 *Level 2: Workflows*

145 A service/capability consists of sequentially and logically interconnected multiple workflows.
146 Workflows are designed to be highly abstracted and modularized for clarity and reconfigurability.
147 Although Workflow has been used to describe the entire DBTL cycle, here we introduce
148 functionally modular workflows for each stage of the DBTL cycle. Table S1 shows 58 biofoundry
149 workflows with short descriptions. Each workflow is assigned to one of the specific Design, Build,
150 Test, or Learn, stages. These workflows encompass the diversity and complexity of synthetic
151 biology experiments, allowing the reconfiguration and reuse of workflows to achieve different
152 functional and executable outcomes. For example, the DNA Oligomer Assembly workflow could
153 be understood to indicate the entire DBTL process for constructing a complete target gene
154 sequence. However, here we use it specifically to define the DNA assembly step where DNA
155 oligomers are assembled. This allows for the development of an ontology of specific actions
156 (workflows) that define the individual steps required to fulfill the entire synthetic biology DBTL
157 cycle. The modularized workflows can be arranged sequentially to perform arbitrary services.
158 Figure S1 represents an example of a protein library construction service.

159

160 *Level 3: Unit-operations*

161 We define unit operations as the lowest abstraction hierarchy level. Unit operations indicate
162 individual experimental or computational tasks. These tasks can be conducted by automated
163 instruments or software tools. By combining unit operations in a sequential manner, workflows
164 can be designed for specific biological tasks. Table S2 and Table S3 show unit operations for
165 hardware and software, respectively. A hardware unit operation can be considered the smallest
166 unit of operation for an experiment corresponding to one or more pieces of equipment. For example,
167 the Liquid Transfer unit operation is an experiment that can be performed by a single liquid
168 handling robot, including PCR setup, dilution, and dispensing. For software unit operations, they
169 are defined based on a software application or package as the smallest unit of operation for an
170 experiment. For example, Protein Structure Generation unit operation is performed for example
171 by RFdiffusion⁷ software application. We propose an initial set of 42 unit operations for hardware
172 (Table S2) and 37 unit operations for software (Table S3). As an example, DNA Oligomer
173 Assembly (WB010) workflow can be represented by 14 unit operations as described in a protocol
174 for synthetic genome synthesis⁸ (Table S4, Figure S2).

175

176 **Flexibility for General Applicability**

177 The modular workflows and unit operations defined here describe various synthetic biology
178 experiments through the reconfiguration and reuse of these elements. However, due to the diversity

179 of biological experiments and the continuous development of improved equipment and software,
180 detailed protocols may vary, which can limit the general applicability of fixed workflows and unit
181 operations. For example, the Liquid Media Cell Culture (WB140) workflow could refer to simple
182 liquid culture for DNA amplification or could include a culture process involving cell-based
183 enzyme assays. In other words, the same workflow or unit operation name can encompass different
184 experimental processes depending on the objectives of the biological experiments. Additionally,
185 workflows or unit operations may differ among laboratories depending on the functionality of their
186 available equipment. For instance, the DNA Extraction (WB045) workflow involves sequential
187 unit operations such as cell lysis and centrifugation. However, some automated equipment can
188 perform the entire DNA purification process in a single operation, so the Nucleic Acid Extraction
189 (UH250) unit operation has been separately added to account for such cases. Similarly, some
190 automated parallel fermenters with functionalities like HT Aerobic Fermentation (UH180) and
191 Microbioreactor Fermentation (UH200) may integrate Microplate Reading (UH370) or simple
192 metabolic/sugar detection functionalities.

193 These challenges highlight the importance of establishing data standards and methodologies for
194 protocol exchange. Existing standards such as SBOL (Synthetic Biology Open Language)⁹ and
195 LabOp (Laboratory Operation Ontology)¹⁰ provide good starting points for describing protocols
196 and workflows in a standardized format. In particular, SBOL's data model is well-suited to
197 represent each stage of the Design, Build, Test, and Learn cycle, and it offers a range of tools¹¹
198 that support data sharing between users, making it compatible with the workflow abstraction
199 proposed in this study. Developing and collecting biofoundry-specific protocols tailored to diverse
200 workflows will be crucial for achieving greater interoperability and reproducibility across
201 biofoundries. This initial version of workflows and unit operations proposed here focuses more on
202 a conceptual framework, definition and classification for biofoundry operations rather than precise
203 definitions. Additionally, a set of unit operations can often resemble familiar protocols with slight
204 variations in methods and naming conventions across laboratories. For example, Golden Gate
205 Assembly, a well-known assembly protocol in synthetic biology, can be viewed as the sequential
206 use of unit operations such as Liquid Handling for DNA part preparation and Thermocycling for
207 enzyme reactions and annealing. This set of unit operations could be named as a distinct Golden
208 Gate Assembly workflow, though further discussions would be required to formalize this
209 classification. However, our proposed conceptual framework allows biofoundry operations to be
210 classified and shared, leading to more standardized operations and the development of calibrants
211 and measurands to allow comparison and interoperability.

212

213 **Software Tools and Data Management**

214 Ensuring that biofoundry-generated protocols and data are reusable, interoperable, and accessible
215 across diverse systems and institutions will require alignment with the FAIR (Findable, Accessible,

216 Interoperable, and Reusable) principles¹², which are essential for effective biofoundry design and
217 software integration. The workflows and unit operations proposed here, for each stage of the
218 DBTL cycle, need to be supported by software tools on multiple levels. For example, the Design
219 step requires CAD (Computer-Aided Design) tools; the Build step requires simulation of
220 laboratory operations and translation of protocols into robotic instructions, via files or application
221 programming interfaces (APIs). The Test stage requires bioinformatics pipelines for data analysis
222 and finally the Learn stage is supported by mathematical and other computational modelling tools.

223 Due to limitations of hardware drivers, a soft integration approach that consolidates data is one of
224 the best options for early-stage biofoundries. Using an integrated database as a single source of
225 truth aligns well with the FAIR principles. However, each unit operation generates a variety of
226 metadata such as operational logs, experimental conditions¹³, and biological raw data^{14,15} requiring
227 careful curation and integration of relevant information. To address this, implementing an API
228 service that runs independently on the computer controlling each piece of equipment, as part of a
229 distributed data management system, would allow seamless accessibility from anywhere.

230 Software tools for biofoundries must efficiently analyze large volumes of biological data and
231 manage a wide variety of diverse experiments. Laboratory Information Management Systems
232 (LIMS) and Electronic Lab Notebooks (ELNs) are essential for comprehensive data management,
233 working in tandem with specialized tools tailored to specific experiments or analytical tasks. Well-
234 known open-source ELN-LIMS solutions include openBIS¹⁶, Aquarium¹⁷, Leaf-LIMS¹⁸ and
235 Galaxy-SynBioCAD¹⁹, while Teselagen Operating System²⁰ and Benchling²¹ are recognized end-
236 to-end commercial solutions. To enable the configurability and flexibility of the workflow
237 approach proposed here, the software tools are best implemented using a modular architecture.
238 This approach accommodates the unique setup of individual biofoundries and makes it easier to
239 add new features or tools to support novel projects. A microservices architecture consisting of
240 smaller, independently functioning applications simplifies adding or modifying services to adapt
241 to specific workflows. This architecture is flexible, scalable, and adaptable to meet diverse
242 biofoundry needs. A microservice architecture with multiple applications specialized for different
243 workflows is more suitable for diverse biofoundry operations than an all-encompassing solution.
244 These applications should be developed with separate front-end and back-end components, adhere
245 to Representational State Transfer (REST) principles²², and be deployed using containerization
246 technologies like Docker and Kubernetes.

247 An example is the Edinburgh Genome Foundry's software suite²³ that enables *in silico* sequence
248 design, modification and cloning; simulation of protocols by modelling microplates and liquid
249 transfers; and QC through design and analysis of sequencing data. The suite is made up of several
250 independent libraries (packages of the Python programming language) that, for each workflow,
251 can be operated individually via a graphical interface (web apps) or are linked together with a shell
252 script. Using scripts to utilize software to perform the required steps, as opposed to a manual
253 procedure, is preferable as it has the same advantages as laboratory automation protocols, namely:

254 batch processing, self-documentation, precision, reproducibility and speed²⁴. Ideally, these tools,
255 and the scripts (which represent protocols), are distributed under a free and open-source license,
256 which is both cost-efficient and allows quick and immediate sharing of expertise and developments
257 between biofoundries and other users.

258 ELNs plays a crucial role in integrating various applications and databases, consolidating the
259 planning and results of experiments, and providing a central source of information. Flexibility can
260 be maximized by using natural language-based software tools, such as electronic lab notebooks,
261 to conduct actual biofoundry experiments. Incorporating natural language to describe experiments
262 enhances the flexibility of workflows and unit operations. A recently proposed approach based on
263 literate programming²⁵ which integrates text and computer code offers new possibilities for future
264 ELN development. The ability to embed computer code in ELNs is crucial for extending their
265 functionality and interacting with other biofoundry applications. In this regard, open-source
266 programming editors like Jupyter notebook, Rstudio(with Quarto), VScode are among the best
267 options for use as a biofoundry ELN. Each of these editors can also be leveraged in cloud
268 environments such as Google Colab, Posit Workbench and GitHub Codespaces, respectively.
269 However, it is important to note that many institutions and companies require their data to remain
270 outside the cloud due to security concerns. Furthermore, as data volumes grow and project
271 durations extend, the high cost of cloud storage can pose a financial burden for biofoundry
272 operations. Therefore, adopting a strategy that combines the advantages of local storage and cloud
273 environments is essential to balance cost and accessibility effectively.

274 For compatibility with ELNs, we illustrate a Tier 3-level Service/Capability example
275 (Supplementary Information) focused on Part DNA Assembly workflows. This example shows
276 the design of workflows (Table S5, Figure S3), provides corresponding experimental records
277 structured according to modular unit operations (Table S6) and its rendered screen shot (Figure
278 S4). Each modular unit operation is documented in Markdown format using natural language, with
279 explicit specifications for title, meta data, inputs, outputs, equipment, reagents, and sample IDs,
280 thereby ensuring full traceability across the workflow. This example illustrates the possibility of
281 how biofoundry experiments built on an abstraction hierarchy framework, can contribute to
282 improved reusability, modularity, and enhanced interoperability across different biofoundries.

283 **Discussions and Future Directions**

284 Compared to a regular laboratory, a biofoundry must comprehensively manage a significantly
285 larger number of equipment, materials, data, experiments, and operations. This necessitates a
286 robust operational framework that ensures seamless functionality, including equipment
287 accessibility, consistent material supply, and rapid analysis of collected data to guide subsequent
288 experimental designs. Biofoundries integrate various automated equipment that should be
289 cohesively connected and substituted with devices from different manufacturers, emphasizing the
290 need for a standardized operational platform. This platform should independently manage user-

291 designed workflows and data, separate from vendor-dependent hardware. RESTful APIs might be
292 useful for effectively translating information exchanged between these workflows and automated
293 equipment. By developing an open lexicon and ontology, multiple public-funded biofoundries can
294 foster cooperation and collaboration on an international scale. While private-sector biofoundries
295 often employ proprietary toolchains that limit broader interoperability, our proposed
296 standardization efforts primarily target public-sector and newly emerging biofoundries that require
297 accessible and flexible operational frameworks. Rather than attempting to encompass all
298 proprietary systems, we emphasize the use of community-driven open-source standards, such as
299 SBOL and LabOp, to overcome technical barriers and accelerate the establishment of interoperable
300 biofoundry infrastructures. A recent report highlighted the need for the development of technical
301 standards and metrics for engineering biology³, and biofoundries could play a leading role in
302 enabling such developments.

303 AI is essential for enhancing the operational efficiency of biofoundries. High construction and
304 operational costs have been identified as significant challenges, with operational expenses
305 particularly threatening the sustainability of biofoundries. AI models capable of analyzing
306 biological and equipment log data generated in biofoundries will be critical for mitigating these
307 risks. The operational efficiency of a biofoundry is directly related to the efficiency of the
308 workflows, such as minimizing consumable usage and saving time and labor within workflows.
309 Optimizing overall biofoundry operations requires a scheduling algorithm that allows multiple
310 workflows to run simultaneously which minimizes interference between them. To optimize the
311 use of limited equipment, it is crucial to continuously monitor the availability of both equipment
312 and materials, maximize the utilization of available time, and effectively coordinate the workflows
313 of various users. AI models are also indispensable for predicting errors and equipment failures
314 during experiments, which helps minimize idle time. This involves collecting data from equipment
315 log files and using additional edge devices to monitor each piece of equipment. Combining AI for
316 real-time task scheduling with predictive modeling for potential failures creates a resilient and
317 adaptive system. Furthermore, biofoundries are uniquely positioned to provide highly curated and
318 quality-assured datasets, which are critical for the development of robust AI/ML models. By
319 leveraging their ability to generate standardized, high-quality data, biofoundries can significantly
320 accelerate advancements in AI/ML-driven research and development. Text-based descriptions of
321 workflows and unit operations in ELNs (Table S6) will be comprehensively extended by large
322 language models, bringing innovative changes to R&D processes in biofoundries.

323 As a follow-up study, developing quantitative metrics to compare workflow performance
324 comparison and evaluate QC is essential for enhancing reproducibility and maintaining high-
325 quality performance in a biofoundry. For example, quality metrics such as cloning success rates
326 can be compared between traditional manual vector construction and automated equipment
327 outcomes. Throughput metrics can measure the workload completed within the same time frame
328 and scale by manual researchers versus automated systems. Capacity metrics could include the
329 number of DNA, plasmids, or RNA synthesized within a given timeframe, as well as the number

330 of strains constructed. Strain construction metrics, often derived from multiple workflows, serve
331 as a representative indicator of overall biofoundry performance. Establishing such metrics requires
332 clear definitions, precise explanations, and measurable formulas. Collaboration within
333 international partner institutions is essential, not only for building workflows but also for gathering
334 input on metric development and selection. Such collaboration will facilitate the identification and
335 adoption of appropriate metrics that accurately reflect biofoundry performance.

336 The abstraction hierarchy framework proposed here will streamline the integration of diverse
337 protocols and serve as a foundation for standardization efforts, ensuring reproducibility and
338 facilitating interoperability across biofoundries. These advancements will enhance the flexibility
339 of workflow management and establish a strong foundation for distributed biofoundry networks.
340 Such networks, supported by AI, standardized data, and workflows, represent a transformative step
341 toward a sustainable bioeconomy and the capacity to address complex global challenges, including
342 pandemics²⁶.

343

344 **Funding**

345

346 This research was funded by the Bio & Medical Technology Development Program (grant number
347 RS-2024-00335300, RS-2024-00508861, RS-2024-00509115, and RS-2024-00445145) of the
348 National Research Foundation, funded by the Ministry of Science and ICT of the Republic of
349 Korea, and the KRIBB Research Initiative Program. PF thanks UKRI EPSRC (EP/5001859/1) and
350 Schmidt Sciences for funding. This work was part of the Agile BioFoundry (agilebiofoundry.org)
351 supported by the U.S. Department of Energy, Energy Efficiency and Renewable Energy,
352 Bioenergy Technologies Office through contract DE-AC02-05CH11231 between Lawrence
353 Berkeley National Laboratory and the U.S. Department of Energy. The views and opinions of the
354 authors expressed herein do not necessarily state or reflect those of the U.S. Government or any
355 agency thereof. Neither the U.S. Government, nor any agency thereof, nor any of their employees
356 makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the
357 accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed
358 or represents that its use would not infringe privately owned rights. The U.S. Government retains
359 and the publisher, by accepting the article for publication, acknowledges that the U.S. Government
360 retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the
361 published form of this manuscript, or allow others to do so, for U.S. Government purposes. The
362 Department of Energy will provide public access to these results of federally sponsored research
363 in accordance with the DOE Public Access Plan ([http://energy.gov/downloads/doe-public-access-](http://energy.gov/downloads/doe-public-access-plan)
364 [plan](http://energy.gov/downloads/doe-public-access-plan)). Funding for open access charge: U.S. Department of Energy.

365

366 **Acknowledgement**

367 We thank all members of the SBLab at KRIBB for their valuable advice and support throughout
368 this research. We also gratefully acknowledge the use of the KRIBB Biofoundry-beta facility.

369

370 **References**

371

372 1. Hillson, N. *et al.* Building a global alliance of biofoundries. *Nature communications* **10**, 2040

373 (2019).

374 2. Crone, M. A. *et al.* A role for Biofoundries in rapid development and validation of automated

375 SARS-CoV-2 clinical diagnostics. *Nature communications* **11**, 4464 (2020).

376 3. Freemont, P. *Engineering Biology Metrics and Technical Standards for the Global*

377 *Bioeconomy*. <http://spiral.imperial.ac.uk/handle/10044/1/110822> (2024) doi:10.25561/110822.

378 4. Martin, H. G. *et al.* Perspectives for self-driving labs in synthetic biology. *Current Opinion in*

379 *Biotechnology* **79**, 102881 (2023).

380 5. Dauparas, J. *et al.* Robust deep learning–based protein sequence design using ProteinMPNN.

381 *Science* **378**, 49–56 (2022).

382 6. Robinson, C. J. *et al.* Rapid prototyping of microbial production strains for the

383 biomanufacture of potential materials monomers. *Metabolic Engineering* **60**, 168–182 (2020).

384 7. Watson, J. L. *et al.* De novo design of protein structure and function with RFdiffusion. *Nature*

385 **620**, 1089–1100 (2023).

386 8. Hutchison, C. A. *et al.* Design and synthesis of a minimal bacterial genome. *Science* **351**,

387 aad6253 (2016).

388 9. Buecherl, L. *et al.* Synthetic biology open language (SBOL) version 3.1.0. *Journal of*

389 *Integrative Bioinformatics* **20**, 20220058 (2023).

390 10. Bartley, B. *et al.* Building an Open Representation for Biological Protocols. *J. Emerg.*

391 *Technol. Comput. Syst.* **19**, 1–21 (2023).

- 392 11. SynBioDex/SBOL-utilities: Command-line utilities and scripts for manipulating SBOL
393 data. <https://github.com/SynBioDex/SBOL-utilities>.
- 394 12. Wilkinson, M. D. *et al.* The FAIR Guiding Principles for scientific data management and
395 stewardship. *Scientific data* **3**, 1–9 (2016).
- 396 13. Barrett, T. *et al.* BioProject and BioSample databases at NCBI: facilitating capture and
397 organization of metadata. *Nucleic acids research* **40**, D57–D63 (2012).
- 398 14. Ham, T. S. *et al.* Design, implementation and practice of JBEI-ICE: an open source
399 biological part registry platform and tools. *Nucleic acids research* **40**, e141–e141 (2012).
- 400 15. McLaughlin, J. A. *et al.* SynBioHub: A Standards-Enabled Design Repository for
401 Synthetic Biology. *ACS Synth. Biol.* **7**, 682–688 (2018).
- 402 16. Barillari, C. *et al.* openBIS ELN-LIMS: an open-source database for academic
403 laboratories. *Bioinformatics* **32**, 638–640 (2016).
- 404 17. Vrana, J. *et al.* Aquarium: open-source laboratory software for design, execution and data
405 management. *Synthetic Biology* **6**, ysab006 (2021).
- 406 18. Craig, T. *et al.* Leaf LIMS: A Flexible Laboratory Information Management System with
407 a Synthetic Biology Focus. *ACS Synth. Biol.* **6**, 2273–2280 (2017).
- 408 19. Hérisson, J. *et al.* The automated Galaxy-SynBioCAD pipeline for synthetic biology
409 design and engineering. *Nature Communications* **13**, 5082 (2022).
- 410 20. TeselaGen Operating System Overview - Teselagen. [https://teselagen.com/teselagen-](https://teselagen.com/teselagen-operating-system-overview/)
411 [operating-system-overview/](https://teselagen.com/teselagen-operating-system-overview/) (2021).
- 412 21. Cloud-based platform for biotech R&D | Benchling. <https://www.benchling.com>.
- 413 22. Fielding, R. T. REST: architectural styles and the design of network-based software
414 architectures. *Doctoral dissertation, University of California* (2000).

- 415 23. Edinburgh Genome Foundry. *GitHub* <https://github.com/Edinburgh-Genome-Foundry>.
- 416 24. Zhang, X. & Jonassen, I. RASflow: an RNA-Seq analysis workflow with Snakemake.
417 *BMC Bioinformatics* **21**, 110 (2020).
- 418 25. Petersen, S. D. *et al.* teemi: An open-source literate programming approach for iterative
419 design-build-test-learn cycles in bioengineering. *PLOS Computational Biology* **20**, e1011929
420 (2024).
- 421 26. Vickers, C. E. & Freemont, P. S. Pandemic preparedness: synthetic biology and publicly
422 funded biofoundries can rapidly accelerate response time. *Nature communications* **13**, 453
423 (2022).

424

425 **Competing Interests**

426 The authors declare the following competing financial interest(s): N.J.H. has financial interests in
427 TeselaGen Biotechnologies and Ansa Biotechnologies.

428

429

430 **Author Contributions**

431 H.K., S.-G.L., and P.S.F. conceptualized the study. H.K., D.-H.L., W.S., and B.H.S. developed
432 the abstraction hierarchy and drafted the manuscript. H.K., N.J.H., and M.S. curated and
433 organized the workflow and unit operation dataset. B.-K.C., D.-M.K., M.-K.O., M.W.C., and Y.-
434 S.J. contributed domain-specific insights into biofoundry operations and synthetic biology
435 protocols. S.J.R., P.V., R.F., R.L.F., and N.S.S. provided critical feedback on interoperability and
436 standardization frameworks. All authors reviewed and edited the manuscript. S.-G.L. and P.S.F.
437 supervised the project and secured funding.

438

439