

Elevate Your FAIR Data Management Strategy With Benchling



Benchling

Executive Summary

Benchling provides cloud-native software solutions for life sciences research and development (R&D). These products are trusted by hundreds of companies across the globe who operate in highly regulated industries, including biopharmaceuticals, vaccines, agricultural products, and industrial biotechnology. These companies are in the midst of a speed-to-innovation revolution, and need comprehensive digitalization of R&D to win the race for the next breakthrough. Benchling has helped over 500 companies achieve successful digital R&D transformations with measurable impacts. Benchling is strongly committed to improving data access, interoperability, and reusability through standardization, and is an active member of Pistoia Alliance and Allotrope Foundation.

This solution brief illustrates how Benchling supports companies in their adoption of FAIR data principles, and demonstrates the value of achieving this transformation in terms of improving productivity, collaboration, data integrity, and decision-making. The intended audience for this document includes scientists, R&D leaders, and R&D information technology (IT) professionals who wish to learn more about how Benchling solutions can be used to improve the findability, accessibility, interoperability, and reusability of their data across their organization. Benchling facilitates each of the four principles in the following ways: by enabling the registration of entities (e.g. cell lines, proteins) with unique ID and persistent HTTP URLs ('Findable'), accessibility of data for authorized users over the web through a UI for humans (HTML) and through APIs for machines ('Accessible'), and annotation with rich metadata using HELM standards, common vocabulary, and ontologies with clear and accessible data usage licenses ('Interoperable' and 'Reusable').

Introduction

Today's life sciences companies have enormous opportunities to transformatively impact human health, agriculture, and energy — from saving lives with breakthrough drugs, to reducing world hunger with lab-grown meat, to protecting the environment through greener products. While these organizations are already making progress on groundbreaking scientific research, it's mission-critical for them to manage and connect complex data and processes, from scientific discovery to commercialization, in order to bring those ideas to market. Digitalization appears to be the best path forward for connecting the entire product life cycle. However, for digitalization to be successful, organizations must adopt informatics solutions that enable the centralization of well-described, standardized, and accessible data, and conform to community standards. FAIR data principles have become an international guideline for high-quality data stewardship on such a platform, and serve as helpful reference points for the development of an effective data management strategy.

FAIR principles articulate specific attributes that data require in order to be optimally reusable by humans and machines. These principles emphasize that all digital objects (data, metadata, etc.) should be “Findable, Accessible, Interoperable and Reusable” both for machines and for people. FAIR was born from an urgent need to improve data management infrastructure, as incomplete, unusable and disorganized data was preventing organizations from extracting the maximum benefit from their investments. A 2018 PwC EU report estimated that a lack of FAIR data costs the European economy €10.2-26 bn1 annually. This high cost was shown to have a detrimental effect on research activities, collaboration, and innovation, due to the significant amount of employee time spent on data searching and collection, as well as unnecessary costs associated with performing redundant studies, and storing duplicate data in multiple siloed applications.

Hence, there is tremendous value in making data FAIR — newly-generated data as well as legacy data. In addition to reducing R&D costs, FAIR data will lead to increased productivity, improved collaboration, enhanced data integrity, and greater scientific and operational insights. Furthermore, it will enable powerful analytical technologies, such as artificial intelligence and machine learning, to automatically access data at scale in order to solve more complex problems. Benchling’s cloud solutions leverage the core principles of FAIR data, and enable life science companies to realize greater value from their R&D data over legacy informatics applications. This solution brief will detail how specific features in Benchling support an organization’s adoption of FAIR data principles, and demonstrate the impact and value of achieving this transformation.

FAIR Principles Drive Effective Data Management

Organization-wide adoption of FAIR data principles requires more than just software. The success of such a program is highly dependent on implementation of the right organizational practices, processes, and mindset for change. To support an organization's FAIR journey, Benchling works to ensure data is captured right from the start, resulting in a solid backbone that can be leveraged to drive more efficient research and development. Benchling's unified platform architecture facilitates robust and reproducible search by natively centralizing all data, and by linking each data point to a useful context (e.g., previous entries, protocols), as well as to associated data points (e.g., related samples, reagents). This centralization and linkage opens up new possibilities for mining data more effectively, driving valuable insights throughout the R&D pipeline.

Benchling's professional services team consists of experts in science and software, who facilitate the smooth transition to FAIR throughout every stage of an organization's journey. These experts work with the key personnel to understand the scientific work being done, craft an appropriate data model using common vocabulary and ontologies, train the relevant teams, and ensure that newly-generated as well as legacy data remain useful throughout the organization.

The following section will detail how Benchling approaches each element of FAIR data principles², to help guide the implementation of an effective data management strategy.



Findability

If a scientist is unable to find data, then it has no value in answering a question or solving a problem. Therefore, data must be structured in a way that makes it easily discoverable and query-able by both humans and machines. This requires effective use of unique IDs, and rich metadata tags.

In Benchling, once an entity is registered with the required metadata, it is assigned a registry ID that serves as a unique and persistent identifier. It also receives its own URL link, which enables anyone with the necessary privileges to access all the information associated with it (F1 and F3). Benchling enforces uniqueness rules, ensuring that two different entities or datasets cannot be mistaken for one another. Furthermore, since each URL link is associated with the unique ID, it remains active even if the data is moved or archived.

F1

(meta)data are assigned a globally unique and eternally persistent identifier

F2

data are described with rich metadata (defined by R1 below)

F3

metadata clearly and explicitly include the identifier of the data it describes

F4

(meta)data are registered or indexed in a searchable resource

Critical metadata is also associated with each registered entity. This metadata describes the entity, and provides useful context (F2). For example, if the entity is an antibody generated in a production lab, then the metadata could include organization-defined attributes as shown in Figure 1:

Plasmid prep used to produce the protein	Transfection protocol used to produce the protein	Cell line used to produce the protein	Name of the person who created the protein
The date when the protein was produced	The location of the protein	The amount of protein produced	Any analytical data associated with the protein (e.g., purity, aggregation, etc.)

Data and metadata are stored in Benchling's cloud application, and can be accessed through the web application, through REST APIs, or through the data warehouse (F4). Benchling's user interface (UI) includes robust, global, Google-like search capabilities, with filtering options that can be used to search through the data using IDs, keywords, target names, metadata, and more. The API provides search and CRUD (Create Read Update Delete) access to almost all data in Benchling through REST endpoints that represent that data in JSON format. The data warehouse, meanwhile, is a read-only SQL database that tracks assay data, biological entities, inventory data, notebook entries, and other records stored in Benchling. The warehouse centralizes an organization's entire research output, facilitating queries that would ordinarily require parsing across multiple data sources.

Figure 1. MAb-002 is the antibody entity that has been registered in Benchling. The entity contains an unique ID (PCo002), URL link, project author, date created, target info, chains info making up the MAb, etc.

The screenshot displays the Benchling interface for the MAb-002 entity. The left sidebar shows the entity's metadata, including its name (MAb-002), registry ID (PCo002), project location, and creation date (8/3/2020 05:09 PM). The main panel shows the entity's details, including its schema (Protein Complex), parent protein complex (MAb-001), protein complex type (MAb), chains (ab2-HV, ab3-LV), binder/chain group (Ab32), target (Target 3), chain(s) (ab2-HV, ab3-LV), plasmid (Plasmid-001), and cell line (CHO-K1).

The right panel shows the experimental results for MAb-002, including a cultivation profile and a KinEsa assay. The cultivation profile table shows data for 10 samples, and the KinEsa table shows data for 2 samples.

Sample ID	Protein Complex	Time (h)	mAb Conc. (mg/L)	Viable Cell Conc. (10 ⁶ Cells/ml)	Glucose Conc. (mmol/L)	Lactate Conc. (mmol/L)	Glutamine Conc. (mmol/L)
1	MAb-002	0	20	0.5	8	23	3
2	MAb-002	10	23	0.8	9	21	2.8
3	MAb-002	20	35	1	11	18	2.2
4	MAb-002	30	40	1.2	15	16	1.5
5	MAb-002	40	48	1.5	18	13	1.1
6	MAb-002	50	56	1.8	19	9	0
7	MAb-002	60	79	2.8	18	8	0.1
8	MAb-002	70	84	3	19.5	7	0.1
9	MAb-002	80	89	2.6	17.5	7	0.1
10	MAb-002	90	90	2.3	18	6.8	0

Sample ID	% active mAb	Kon	Koff	Kd	Signal 100%	% Error	
1	MAb-002	84	1.2	12	4.2	1.23	0
2	MAb-002	45	12	1.2	12	1.36	0



Accessibility

For data to be accessible, it needs to be easily searchable by humans and computational systems, without requiring extensive knowledge on its creator or date of creation. Benchling is SaaS-based, and can be accessed by authorized personnel using the unique identifier and HTTP URL (internet link), via any standard computer with an internet connection (A1 and A1.1). Benchling also conforms to the openAPI standard, which allows computational systems to discover and understand the capabilities of the service without access to source code or documentation, and with a minimal amount of implementation logic (A1.1).

Accessibility does not imply being open, instead it means that only authorized users can interact with the system. Individual users can be configured using a granular permission structure, which offers flexible customer-admin defined user roles, along with the ability to grant specific types of access to external users (A1.2) using the UI (Figure 2) and API (Figure 3).

A1

(meta)data are retrievable by their identifier using a standardized communications protocol

A1.1

the protocol is open, free, and universally implementable

A1.2

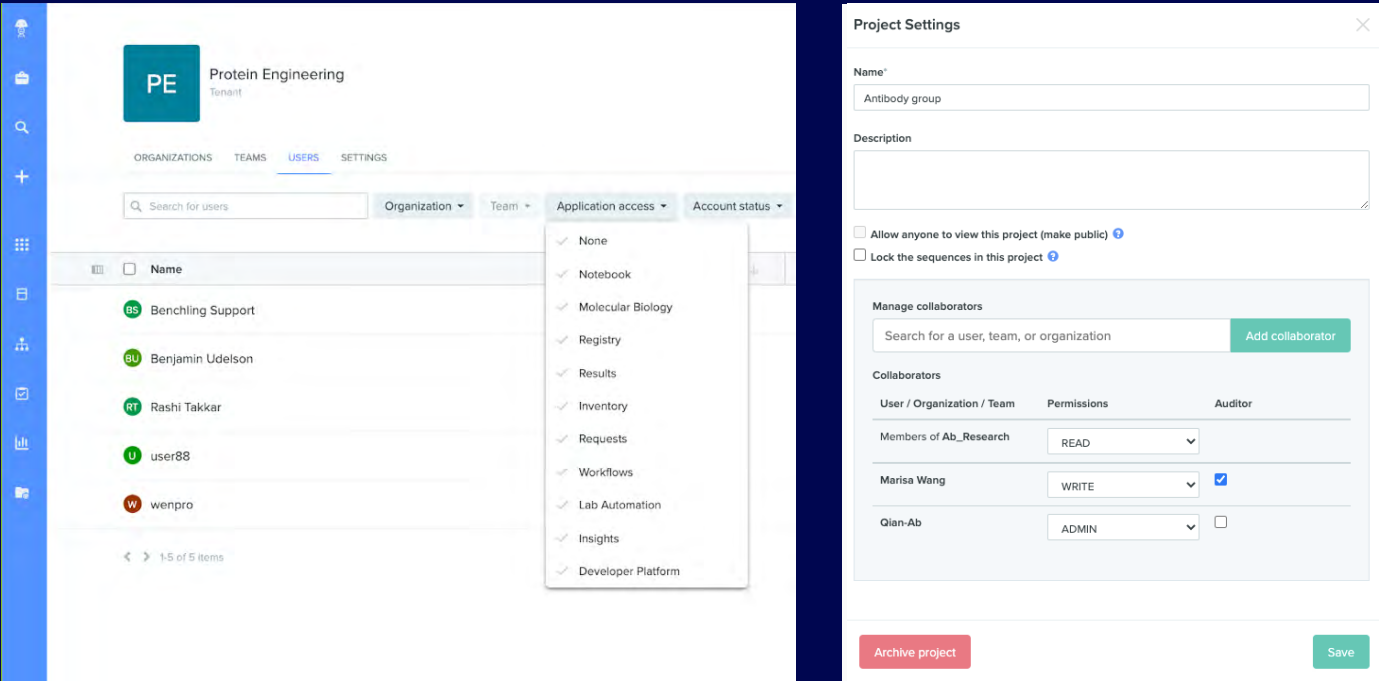
the protocol allows for an authentication and authorization procedure, where necessary

A2

metadata are accessible, even when the data are no longer available

User access controls can be integrated with an organization’s existing single sign-on (SSO) and multi factor authentication (MFA) protocols, ensuring that only authorized individuals can access and make changes to the system — further enhancing security and mitigating risk. For organizations that desire an additional layer of security, Benchling also provides rich APIs that can be integrated with existing enterprise security systems. These REST APIs communicate with the client by automatically using the HTTPS protocol for authentication and protection of data in transit (A1, A1.1, A1.2). All data and communication within Benchling is protected at rest using AES 256-bit encryption, and in transit using Transport Layer Security (TLS) encryption 1.2 or higher.

Figure 2. Customer admins can authorize application(s) access and grant privileges (read, write, admin) to individual users directly in the UI.



Raw and formatted data, as well as the metadata, are stored in the cloud, and are accessible through the UI, or through automated APIs, even if the original data has been archived (A2). This allows other scientists in the organization to access the metadata, which are valuable for planning research — especially replication studies.

Figure 3. API credentials can be used to access, read, and write data to Benchling with an API key unique to users.

The screenshot displays the Benchling API documentation interface. On the left is a sidebar titled 'Resources' with a list of categories: AA Sequences, Assay Results, Assay Runs, Batches, Blobs, Boxes, Containers, Custom Entities, DNA Alignments, DNA Oligos, DNA Sequences (highlighted in blue), Dropdowns, Entries, Events, Exports, Folders, Inventory, Lab Automation, Label Templates, Locations, Oligos, Plates, Printers, Projects, Registry, Requests, Schemas, Tasks, Warehouse, and Workflows. The main content area is divided into two sections. The top section, 'DNA Oligos', includes a description: 'DNA Oligos are short linear DNA sequences that can be attached as primers to full DNA sequences. Just like other entities, they support schemas, tags, and aliases.' and a right-pointing chevron. The bottom section, 'DNA Sequences', includes a description: 'DNA sequences are the bread and butter of the Benchling Molecular Biology suite. On Benchling, these are comprised of a string of nucleotides and collections of other attributes, such as annotations and primers.' and a downward-pointing chevron. Below the 'DNA Sequences' description is a list of API endpoints, each with a colored button indicating the HTTP method (blue for GET, green for POST, purple for PATCH), the endpoint path, and a brief description of the action. The endpoints are: GET /dna-sequences (List DNA sequences), POST /dna-sequences (Create a DNA sequence), GET /dna-sequences/{dna_sequence_id} (Get a DNA sequence), PATCH /dna-sequences/{dna_sequence_id} (Update a DNA sequence), POST /dna-sequences:archive (Archive DNA sequences), POST /dna-sequences:autofill-parts (Autofill DNA sequence parts), POST /dna-sequences:autofill-translations (Autofill DNA sequence translations), POST /dna-sequences:bulk-create (Bulk Create DNA sequences), GET /dna-sequences:bulk-get (Bulk get DNA sequences by ID), POST /dna-sequences:bulk-update (Bulk Update DNA sequences), and POST /dna-sequences:unarchive (Unarchive DNA sequences). At the bottom of the main content area is a section titled 'Dropdowns'.

Resources

- AA Sequences
- Assay Results
- Assay Runs
- Batches
- Blobs
- Boxes
- Containers
- Custom Entities
- DNA Alignments
- DNA Oligos
- DNA Sequences**
- Dropdowns
- Entries
- Events
- Exports
- Folders
- Inventory
- Lab Automation
- Label Templates
- Locations
- Oligos
- Plates
- Printers
- Projects
- Registry
- Requests
- Schemas
- Tasks
- Warehouse
- Workflows

DNA Oligos

DNA Oligos are short linear DNA sequences that can be attached as primers to full DNA sequences. Just like other entities, they support schemas, tags, and aliases. >

DNA Sequences

DNA sequences are the bread and butter of the Benchling Molecular Biology suite. On Benchling, these are comprised of a string of nucleotides and collections of other attributes, such as annotations and primers. v

GET	/dna-sequences	List DNA sequences
POST	/dna-sequences	Create a DNA sequence
GET	/dna-sequences/{dna_sequence_id}	Get a DNA sequence
PATCH	/dna-sequences/{dna_sequence_id}	Update a DNA sequence
POST	/dna-sequences:archive	Archive DNA sequences
POST	/dna-sequences:autofill-parts	Autofill DNA sequence parts
POST	/dna-sequences:autofill-translations	Autofill DNA sequence translations
POST	/dna-sequences:bulk-create	Bulk Create DNA sequences
GET	/dna-sequences:bulk-get	Bulk get DNA sequences by ID
POST	/dna-sequences:bulk-update	Bulk Update DNA sequences
POST	/dna-sequences:unarchive	Unarchive DNA sequences

Dropdowns



Interoperability

To be considered interoperable, data must be able to be integrated with other data sources, in order to enhance knowledge discovery and drive informed decision-making. To achieve interoperability, an informatics system must use broadly applicable vocabulary and ontologies. Benchling's configurable data model is built with standard vocabulary and ontologies that are agreed upon prior to implementation, and can map to any scientific process. While Benchling encourages the use of industry standards that follow FAIR principles, customers are free to use their own vocabularies, in order to conform to standards applicable to their industry (I1 and I2). Figure 4, below, shows an example of a recommended data model for protein engineering/production teams, which Benchling has developed based on hundreds of implementations. The data and associated metadata can be extracted using the Benchling API, and can then be processed by machines.

I1

(meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.

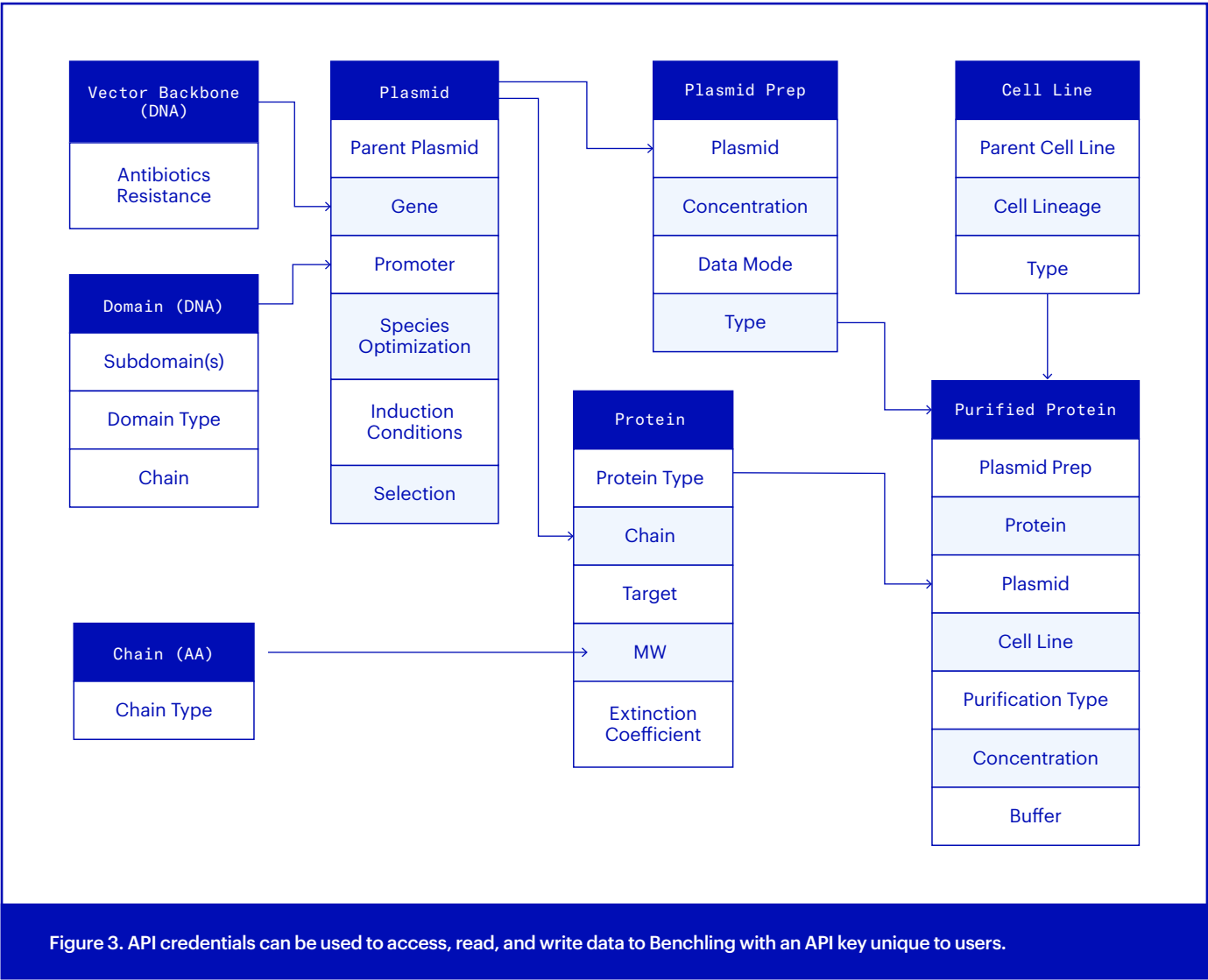
I2

(meta)data use vocabularies that follow FAIR principles

I3

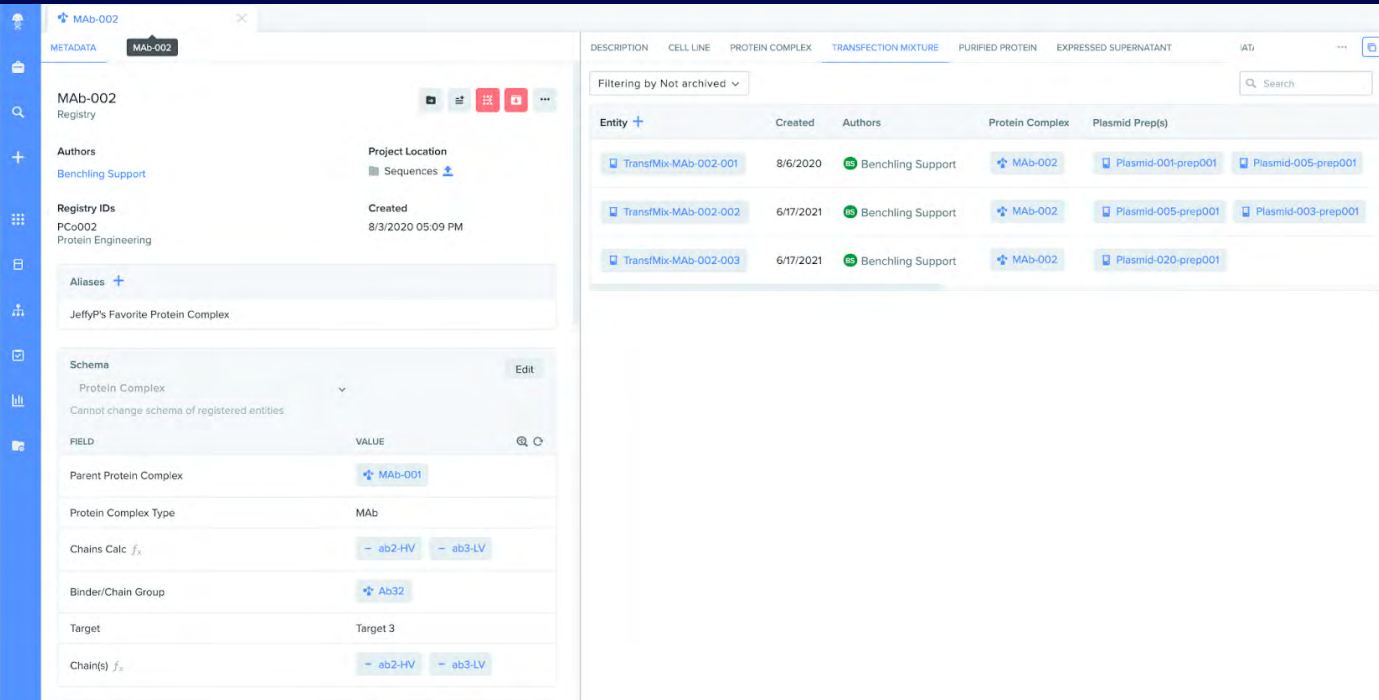
(meta)data include qualified references to other (meta)data

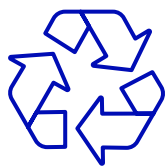
To aid interoperability between different systems, Benchling also imports standard IDs, such as the Ensembl ID, NCBI Accession and relevant metadata when a DNA or an amino acid sequence is imported into the system from an external database.



In Benchling, meaningful “smart links” are created between data and metadata resources, enriching researchers’ contextual knowledge about the data. In addition, for multi-stage workflows, Benchling enables the creation of connections between “child” entities created from previous “parental” entities as seen in Figure 5. For example, “child” entities are typically lots, dilutions, or preps of a specific “parental” entity, which require unique metadata such as prep type, date created, creator name, concentration, etc (13).

Figure 5. TransfMix-MAB-002-001, ...-002, ...-003 are “child” entities (right) are multiple preps of the parental MAB-002 antibody represented on the right





Reusability

When existing data can be leveraged to answer new questions, it becomes reusable. Benchling's unified informatics platform makes data points reusable, by making it easy for organizations to link them to the context under which they were originally generated — such as the introduction, materials used, protocol used, date of generation, and experimental parameters, with clear template versioning as seen on Figure 6. Experiment results, for instance, can automatically be linked to all of their associated data throughout the platform, including previous lab notebook entries, related samples, reagent preps, and DNA sequences. Each piece of data is natively linked with its context; users can click on any single element and instantly view the whole picture (R1 and R1.2). This also allows easy consolidation of data across multiple experiments, resulting in deeper insights (Figure 7).

Benchling enables the defining, storing, computing and comparing biologics in a standard manner, so information can easily be shared among scientists and organizations. Benchling has adopted the HELM standard to deal with many subtleties of biologics, such as post-translational modifications, cross linking of domains, linkers and ambiguity (R1.3). Benchling also supports the FASTA format for representation of DNA and amino acid sequences (R1.3).

R1

meta(data) are richly described with a plurality of accurate and relevant attributes

R1.1

(meta)data are released with a clear and accessible data usage license

R1.2

(meta)data are associated with detailed provenance

R1.3

(meta)data meet domain-relevant community standards

Additionally, Benchling promotes reusability by allowing data to be released with a clear and accessible usage license, promoting full transparency and detailing the conditions under which the data can be used in the Customer Agreement (R1.1).

Figure 6. The protocol, template with clear versioning, project and the entity being purified are connected using “smart links”.

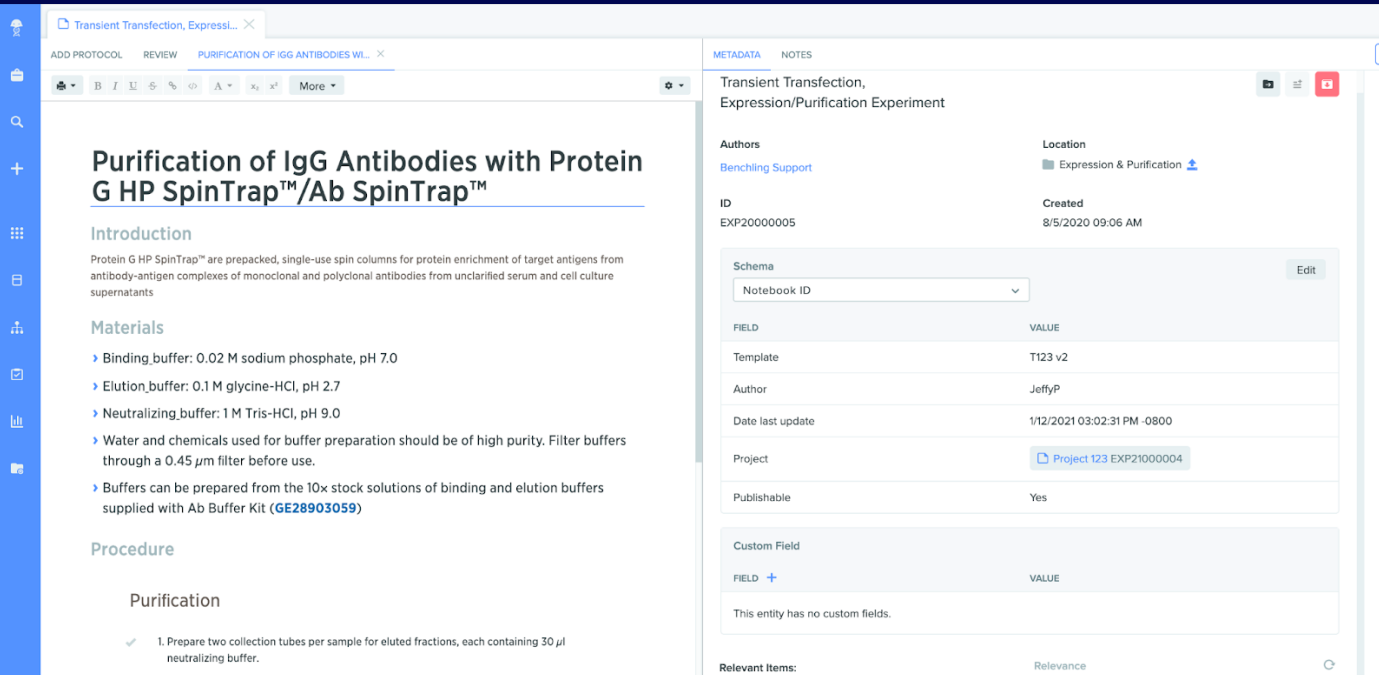
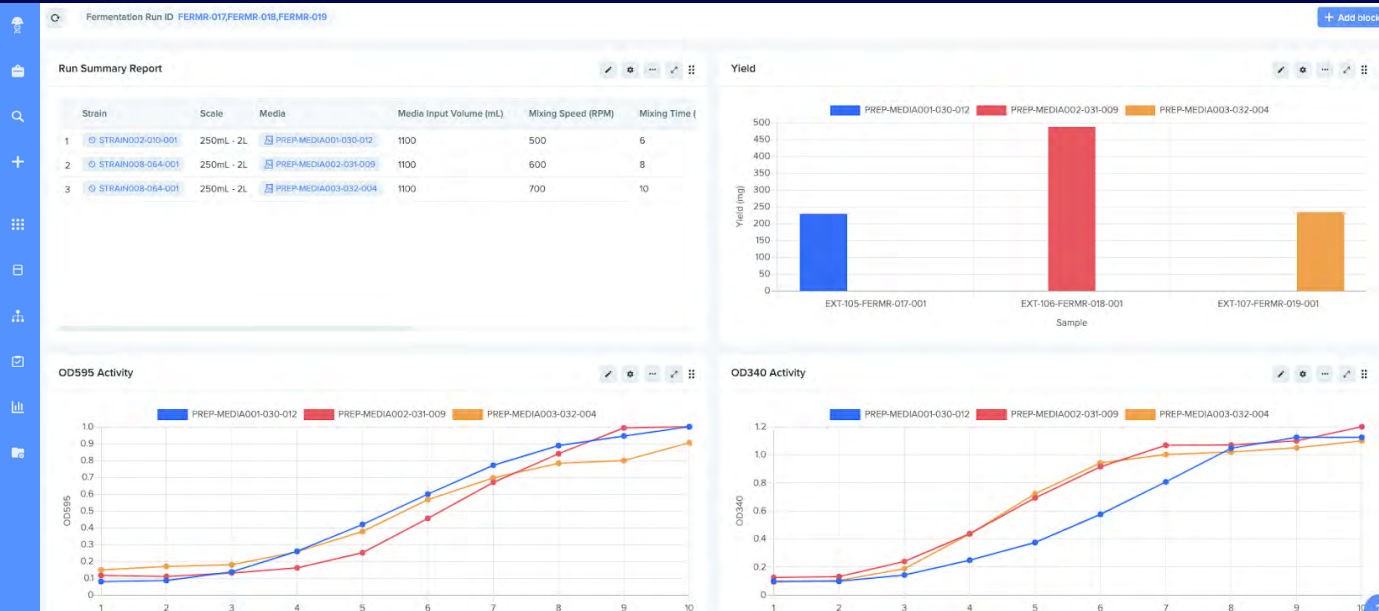


Figure 7. Three fermentation runs across multiple weeks, and by different users can be easily consolidated to generate insights.



Security and Privacy

Although security and privacy are not included in the FAIR principles, they are mission-critical to any effective data management strategy. Benchling maintains the highest levels of security, and compliance, and is committed to protecting its customers' privacy rights. The company has aligned its practices with Privacy Shield, GDPR and CCPA regulations, and is ISO 27001 certified.

Benchling utilizes Amazon Web Services (AWS) for hosting. AWS utilizes world-class infrastructure security practices, and maintains multiple security and compliance certifications, including ISO 27001, SOC 1, and SOC 2. Within Benchling's AWS cloud environments, customer data is stored in S3 buckets, which provide granular, auditable control over data protection, along with programmatic verification of data security. Furthermore, Benchling's high-redundancy data storage practices minimize the risk of data loss. Structured data is stored in Postgres databases, which are configured to encrypted daily and weekly backups with synchronous replication . All content stored on Benchling's system is retained throughout the entire lifetime of each customer. Daily backups are kept for 35 days or more, to allow for granular data restoration, while weekly backups are kept for at least a year.

FAIR Principles Drive Effective Data Management

As the above sections demonstrate, FAIR principles can increase productivity, improve collaboration, enhance data integrity, and bolster scientific and operational decision-making. These improvements deliver measurable business value for stakeholders, helping a scientific organization chart a strategic path to success.

The following tables summarize a number of quantitative benefits Benchling’s customers report as a result of adopting FAIR data principles.

Time Metrics	Benchling Impact in Time Reduction
Plasmid prep used to produce the protein	53%
Time spent finding a physical location of a sample(s)	66%
Time spent searching for all data associated with a specific sample	67%
Time spent compiling data on a specific set of samples for analysis and reporting purposes	69%
Time spent on collaboration i.e. request submission and status tracking	63%
Time spent repeating experiments due to lost, unclear or incomplete data	45%
Time spent preparing for or reviewing data in meetings with your peers	60%

Satisfaction / Usage Metrics	Benchling Impact – Improvement in Satisfaction
Level of satisfaction with data capture and reporting tools and systems (e.g., disparate tools before; Benchling now)	53%
Level of confidence in data quality	72%
Level of ease in finding previous versions of experimental protocols	77%
Level of ease in collaborating and transitioning experiment / sample data from one group to the next group in a scientific workflow	71%
Level of ease in finding the physical location of a sample and reagents	52%
Level of ease in tracking reagents and their properties (metadata)	125%
Level of ease in finding data or protocols owned by a previous team member	51%

458

Hours saved per year with Benchling

\$34,381

FTE cost saved per scientist per year with Benchling

>\$3.4m

Cost-savings for an organization with 100 scientists

All teams surveyed for this analysis agreed that the adoption of FAIR data with Benchling resulted in the following improvements:

Centralized and standardized data collection and management within a unified data model	Transparent, traceable and searchable sample and experimental data, including metadata	Rapid and accurate data sharing and collaboration across teams, and with broader stakeholders
Improved scientist productivity and efficiency, and reduced time spent on logistical tasks	Optimized and automated scientific workflows with end-to-end experimental enhancement	

Overall, organizations report that FAIR data with Benchling delivered significant near-term value: 62% productivity gains, 71% improved cross-team collaboration, 72% higher data integrity, and ~\$3.4M yearly decreased R&D costs for a 100-scientist organization. Many executives also reported that the Benchling’s cumulative value across all these dimensions helps them reach milestones quickly, and get the right product into the market.

References

¹ Open Data Policy and Science Cloud. Cost-Benefit analysis for FAIR research data — Cost of not having FAIR research data. (2019). European Commission. <https://op.europa.eu/en/publication-detail/-/publication/d375368c-1a0a-11e9-8d04-01aa75ed71a1/language-en>

² Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016). <https://doi.org/10.1038/sdata.2016.18>

Conclusion

FAIR data is critical to solve the inefficiencies that have arisen from legacy data management practices, such as data silos, inconsistent terminologies and lack of sufficient context (metadata) on the data. Benchling is designed from the ground up to support FAIR data. Benchling facilitates each of the four principles in the following ways: by enabling the registration of entities (e.g. cell lines, proteins) with unique ID and persistent HTTP URLs ('Findable'), accessibility of data over the web through a UI for humans (HTML) and through APIs for machines ('Accessible'), and annotation with rich metadata using community standards, formats and ontologies ('Interoperable' and 'Reusable'). Benchling has provided significant benefits for hundreds of customers in improving FTE productivity, facilitating collaboration and reducing costs. Adopting the Benchling platform will enable the creation of a centralized and unified data foundation that is necessary for tomorrow's leading-edge life science discoveries.

Join over 300,000 scientists using Benchling to power life science R&D.

[Request a Demo](#)