



Best Practices for Managing Your Team in Domino

This brief presents concrete tips for using Domino to manage your team more effectively.

Introduction

Well-managed data science teams deliver more models and insights faster. With the right practices and systems in place, you can speed up onboarding for new hires, mitigate key-man risk, reduce wasted effort re-inventing the wheel, as well as spot problems and opportunities faster.

To read more about management best practices, please read our [Practical Guide to Managing Data Science at Scale](#).

How Domino Helps

In addition to increasing productivity for data scientists by providing modern tooling that helps attract talent and reduce turnover, Domino makes it easy to:

- 1. Onboard new data scientists faster** by eliminating the time they spend configuring environments, finding data and other reusable assets, and getting context on past work.
- 2. Keep your finger on the pulse of in-flight work** without disrupting people's research flow.
- 3. Create standards and templates** for how work gets done, so you can ensure consistent application of best practices across your team.

These techniques are often a challenge because data scientists work in siloed environments, such as their laptops or desktops. Centralizing data science work is the key to managing it more effectively. This key insight is the core of Domino's solution.

How it Works in Domino

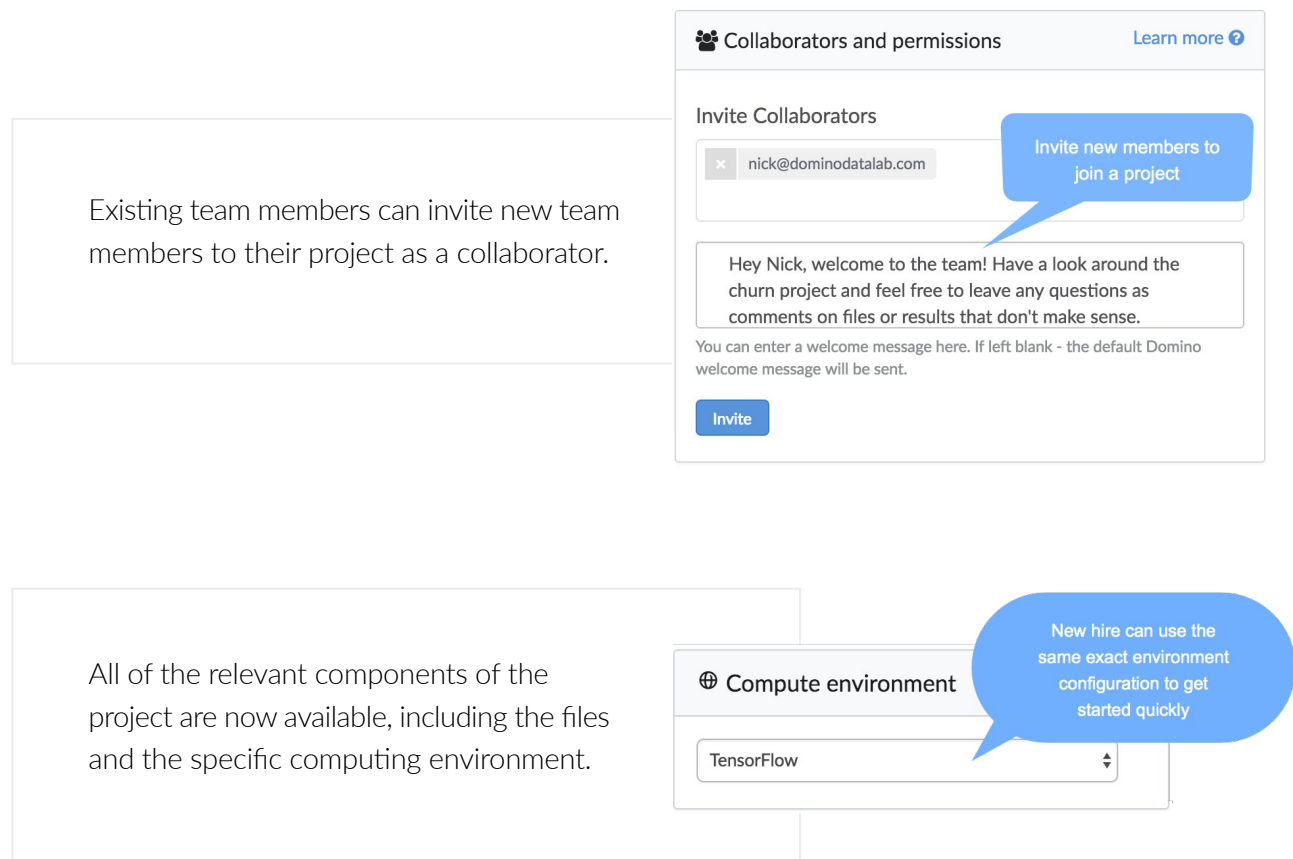
Domino offers a number of features that help data science managers better manage their growing teams. This guide provides a few best practices based on our experience with large and growing data science teams. We delve into best practices for onboarding, monitoring in-flight work, and creating standardized templates.

Onboarding new data scientists

We see two common workflows when new hires join a team: being added to an in-flight project and being asked to start a new project.

1. New hires are added to an in-flight project.

Without a centralized platform, they would waste time getting their environment setup correctly and learning the context of what's already happened in the project. With Domino, both of these barriers dissolve.



The new team member can review the Runs dashboard to see a summary of all the work that has been done, including key metrics and any discussions.

The screenshot shows the Domino Runs dashboard. On the left, a list of runs is displayed, including 'Testing new sentiment feature', 'Cleaning the call transcript data' (highlighted), 'Initial tests of API connection to c...', and 'Preliminary EDA of customer data'. The main panel shows the details for the 'Cleaning the call transcript data' run, which is titled 'Cleaning the call transcript d...'. It includes a 'Discussion' tab with comments from users like 'mac' and 'tim'. A blue callout bubble points to the discussion area, stating: 'Context and discussion are preserved to come up to speed fast'.

2. New hires are asked to start a new project.

Without a central platform, they'll waste time tracking down info on how to access different internal data sources, and how to use the right set of tools for the project.

Domino lets users create canonical data sets which preserve connection details, query details, file revisions, and any prep or cleaning scripts.

The screenshot shows the 'Data Sets' catalog in Domino. It lists various data sets with their paths and associated tags. The 'mac/clickstream_data' entry is highlighted with an orange box. A blue callout bubble points to this entry, stating: 'Catalog datasets along with processing scripts'. Other data sets listed include 'mac/census_data', 'mac/call_transcripts', and 'mac/customer_data'.

These intermediate artifacts can then be imported into downstream projects for more exploration, additional feature engineering, or formal model development.

Project	Location	Release	Imports	View	Remove
mac/clickstream_data	/mnt/mac/clickstream_data	Latest	Files	View	Remove

New users select the appropriate set of tools from pre-existing environments or configure their own without impacting other users on shared servers.

\$ Compute environment

Global
 Default
 Default + pluggable notebooks
 GPflow + tensorflow
 Intel Python
JupyterLab
 Kaggle showcases
 Py3/TF/cuDNN
 Python 2.7 + CUDA 8
 Python 3.6
 Python showcases
 R showcases (including shiny support)
 Zeppelin 0.72 with Spark 2.1.0
 [Domino Managed] Default 2017-02
 [Domino Managed] GPU Tools
 [Domino Managed] Anaconda Python 2.7
 [Domino Managed] Python 3

Easily toggle between desired tool set for new project

Keeping your finger on the pulse of the team

Domino lets managers easily check in on the progress of their team to provide guidance and coaching regardless of the tools they use. Since all of the work takes place in a centralized platform, managers can view all ongoing projects in the Projects Overview and see what's being actively worked on.

Projects

[Your Projects](#) 7
 [Projects You Collaborate On](#) 3
 [Suggested Projects](#) 5
 [Popular Projects](#) 5

jonathanmaher / credit_risk_d...

1

All good!

Medium

Running

Previous Runs

Hardware

Latest results 4 minutes ago

Just getting started

johnconway / churn_predictor

Model to predict retail customer churn and recommend optimal intervention strategy

Machine Learning Model Selection endpoint

evaluate model validation

49d 3h 1m

All good!

Default

Total time

Previous Runs

Hardware

Latest results 8 hours ago

Exporting files

Train Model

johnjoo / support_sentiment

A model to analyze the call center customer interaction sentiment based on recorded transcripts sampled over the last year

Machine Learning

model validation

49m

1 Failed

Default

Total time

Previous Runs

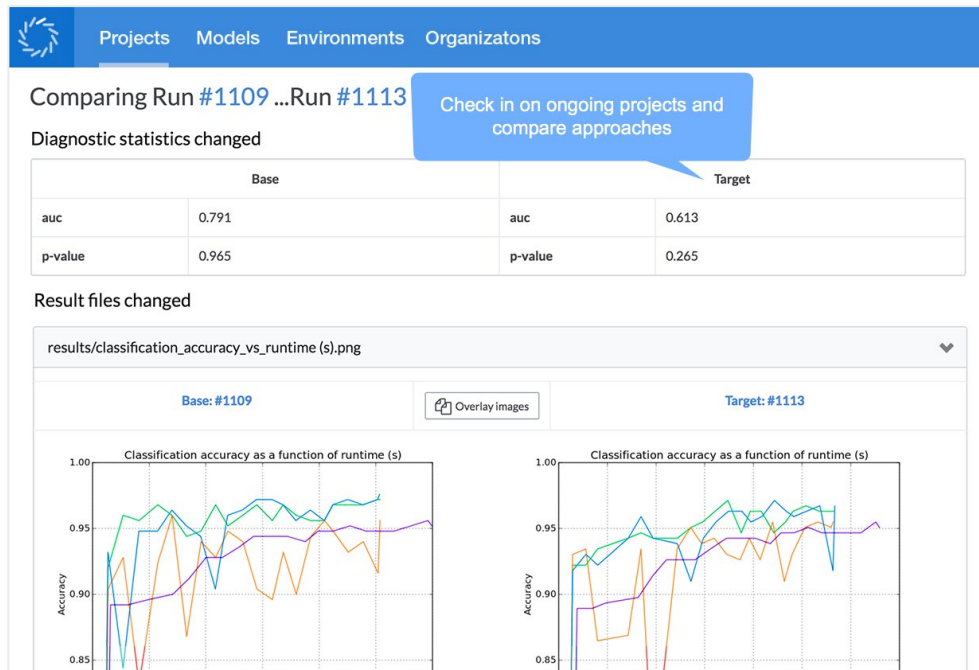
Hardware

Latest results 6 days ago

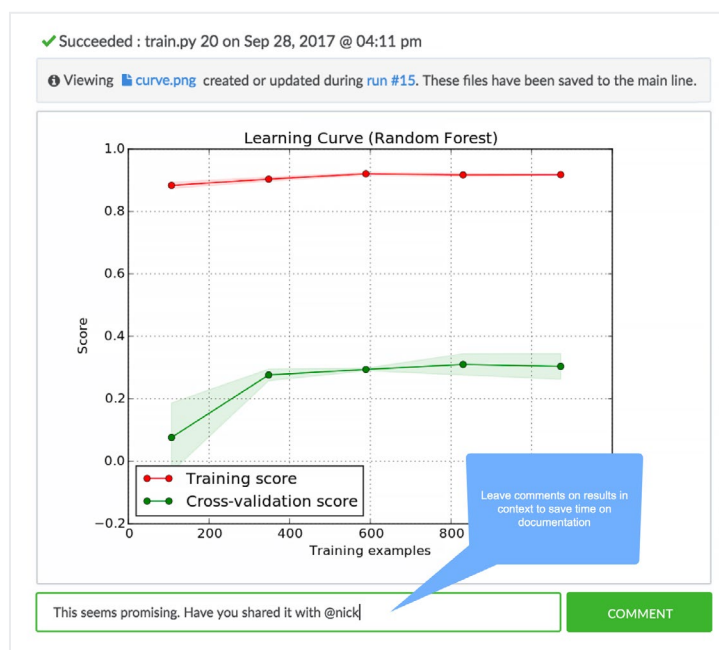
Exporting package

Evaluate Model

If managers want more details or something looks amiss, they can dive into notebooks and leave comments on results; all without disrupting their team's flow. If more significant coaching is necessary, they can run their own experiments within the same project and compare results to help explain the impact of any changes.



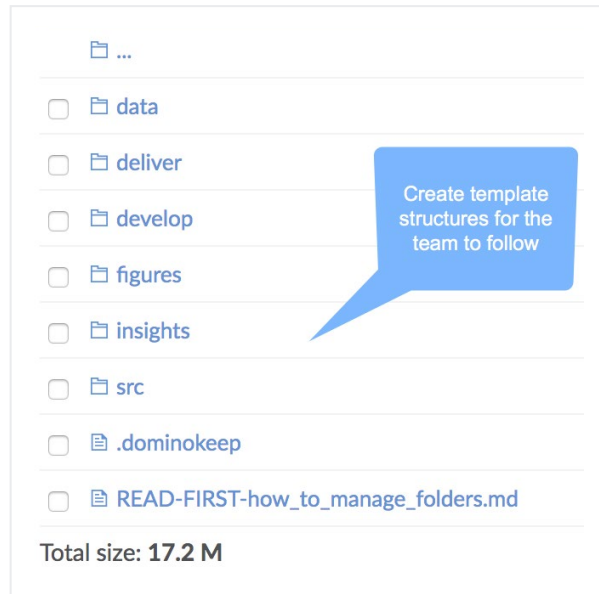
Many teams use Domino as the basis for their Agile stand-ups; showcasing the work each data scientist has done in a sprint on a shared screen. Teams record comments or concerns directly in the project to preserve the context of the feedback.



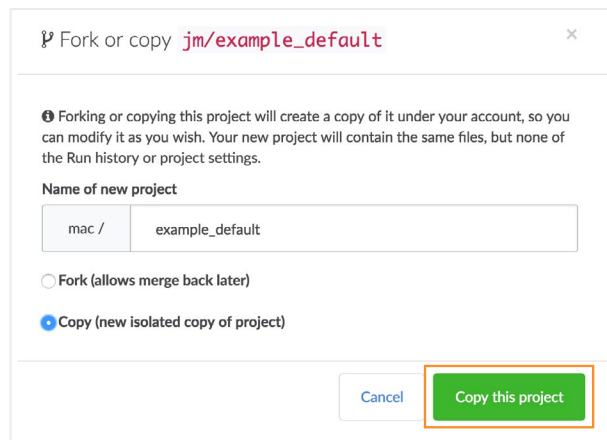
Provide templates of best practices (documentation, file structure, tool usage, etc.)

Many managers struggle to encourage or enforce certain practices around documentation. This struggle leads to headache in collaboration as different peoples' systems conflict. Domino allows teams to create template projects which are available to all members; to showcase best practices across a number of research steps.

For example, if a team has a canonical file structure for each new project



they can extend this to include a template for a model requirements document, standard data dictionaries, or stakeholder presentations. Any changes to templates are automatically revised by Domino. Each new project copies this template, dramatically reducing the time to come up to speed on a new project and ensuring everyone knows where to look for key artifacts.



This also ensures that if someone leaves the team, their work is immutably stored on Domino and easily navigable by new team members.

Best Practices for People and Process Changes

Data science management best practices in Domino can extend all the way to the data science hiring process. Multiple Domino customers use Domino as their candidate assessment portal. Specifically, data science managers temporarily invite candidates into their Domino deployment to assess their candidate's abilities. The candidate is added to an existing project that is pre-loaded with non-sensitive data, and a standardized environment. This enables the candidate to focus on model development and delivery versus infrastructure configuration. With this functionality, hiring managers showcase the organization's embrace of modern tooling which is critical for attracting and retaining top talent.

Stepping back, managing a data science team well requires more than just new technology. Our most successful customers build robust training programs, bring data scientists together across departments for lunch-and-learns, plus establish [data science-specific KPIs](#). Moreover, managers build a diverse team with a variety of backgrounds to take on the ever-increasing variety of projects in their purview. See [our guide](#) for a more detailed discussion on implementing best practices based on the experience of leading organizations like Allstate, Monsanto, and Moody's Analytics.

Conclusion

The practice of data science management is increasingly separating good teams from great teams. A number of customers use the Domino platform as a way to better manage their growing team, from the day someone is hired to the moment they transition to a new opportunity.



About Domino Data Lab

Domino Data Lab provides the world's most advanced data science platform, powering organizations that are using predictive models to drive their business. Companies including Allstate, Coatue, Mashable, Monsanto and startups alike use Domino to accelerate breakthrough research, increase collaboration and productivity of data scientists, and more rapidly deliver models to drive business impact.

This content is copyrighted © 2017 by Domino Data Lab, Inc. and cannot be reproduced or shared without prior expressed written permission. Domino Data Lab and the Domino Data Lab logo are registered trademarks of Domino Data Lab Inc.