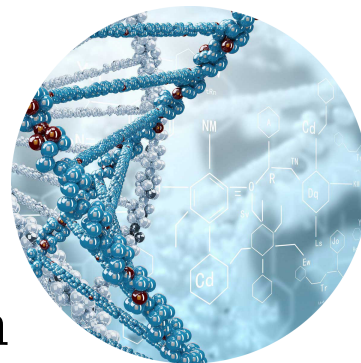


Manual for the Computational Data Science Lab at the University of Texas Arlington



This manual summarizes a set of resources, guidelines, and rules for successful education, research, and communication of Science among the members of the Computational Data Science (CDS) Lab in the College of Science at The University of Texas at Austin.

For more information please contact:

[Amir Shahmoradi](#) ,
Principal Investigator (PI), CDS Lab,
Assistant Professor, Department of Physics,
Data Science Program, College of Science,
The University of Texas at Arlington
655 W Mitchell St | SEIR 365
Arlington, TX 76010-1765
a.shahmoradi@uta.edu
Tel: (817) 272 9047
shahmoradi.org
cdslab.org

Contents

0.1	Introduction	2
0.2	Current and Past Members	3
0.3	Communication	3
0.3.1	Instant Communication	3
0.3.2	Other Communications	3
0.4	Introduction	4
1	Meetings	5
1.1	1st Joint Meeting: Monday, 10:00-11:15, Sep 10 2018	5

0.1 Introduction

The Computational Data Science (CDS) Lab at The University of Texas at Arlington (UTA) carries out research in a wide range of scientific disciplines, including Biomedicine, Biophysics, Astrophysics, Mathematical Modeling, and Scientific Software Development, all of which require intensive usage and development of Computational and Data Science methodologies and algorithms.

While the research carried out in the CDS Lab is purely theoretical / computational, we collaborate frequently and welcome new collaborations with experimental groups and clinicians across the UTA campus, DFW Metroplex, as well as nationally / internationally.

The CDS Lab is proudly a member of the College of Science at The University of Texas at Arlington and is heavily involved in the development of the Data Science Program as part of the UTA's Strategic Plan | 2020.

This manual is written for the current, new, or prospective members of the CDS Lab to help them communicate, collaborate, and carry out research at the highest levels of efficiency and scientific integrity possible.

0.2 Current and Past Members

For information about current and past members, as well as the Lab PI, please see <https://cdslab.org/people/>.

0.3 Communication

0.3.1 Instant Communication

- **Slack:** <https://cdslaborg.slack.com>

Slack is the primary tool for instant communication among members of the CDS Lab. All lab members should download the Slack Application on their mobile device, and use it for instant communication with individual or all members of the lab, including the PI (Amir). Non-urgent lengthy time-consuming issues can be discussed via email or in person.

0.3.2 Other Communications

The CDS Lab has webpages on multiple media platforms, including,

- **Medium:** <https://medium.com/cdslab>

This is a social platform that we use mostly for communicating medium-to-long pieces of scientific research or educational material, performed by the lab members, in the form of blog-posts with the world. All lab members are welcome and encouraged to contribute posts to this page. Medium is a highly-regarded platform for advertising your work to the entire world. Writing an elegant readable article on this page that could attract public's attention, could significantly boost your profile as an expert in your field.

- **Facebook:** <https://fb.me/cdslab>

Facebook is mostly used for communicating the lab news and articles with the Lab members as well as public community on Facebook. You can follow the page by clicking on the link provided above.

- **Twitter:** <https://twitter.com/cdslaborg>

Similar to Facebook, Twitter is mostly used for communicating the lab news and articles with the Lab members as well as the public community on Twitter. You can follow the page by clicking on the link provided above.

- **Instagram:** <https://www.instagram.com/cdslaborg>

Similar to Facebook and Twitter, Instagram is mostly used for communicating the lab news and articles with the Lab members as well as the public community on Instagram. Obviously, any news that appears on Instagram has to be followed by an illustration or photo that would serve as a visual explanation of the news.

- **GitHub:** <https://github.com/cdslaborg>

GitHub is a highly popular code repository that we use for keeping track of changes in our scientific codes and data as well as for sharing our codes and data processing pipelines with the entire world, permanently, once it is complete. As a member of a data-intensive lab, you are expected to have

a GitHub account. There are also other, almost equally-capable online code repositories available, such as Bitbucket. However, at the moment GitHub is the primary code repository of the lab.

- **Dropbox:** <https://dropbox.com/>

Like GitHub, we often use Dropbox as way of communicating all digital information such as code, data, articles, manuscripts, but only for our own internal usage (unlike Git and GitHub which are used for professional version controlling of the codes, as well as for sharing scientific research pipelines and information with the entire world. If you don't have a Dropbox account yet, please let Amir know as soon as possible to send you an invitation email.

0.4 Introduction

Chapter 1

Meetings

The names of the attendants who suggest or advocate specific ideas in each meeting follows (in boldface font) the description of the idea itself.

1.1 1st Joint Meeting: Monday, 10:00-11:15, Sep 10 2018

Attendees (alphabetical order):

Amir Shahmoradi , **Amir Farbin** , **Leili Shahriyari** , **Daniel Welling**

Agenda:

- Discuss the development of Data 2401: Intro Scientific Computation course and other Data courses in the broader scope.

List of topics highlighted / discussed:

- **“Intro Data Science” Curriculum Development:**
 - **Li Wang** from the Mathematics department has already developed the course material for Intro Data Science and is currently teaching the course.
 - **Course structure:** **Li Wang** is using Python instead of Excel (as in the original plan) for this course. This may be a good idea since Excel is rarely used a Data Science software tool in the community (**Leili Shahriyari**). However, to ensure the original goals of the course are met, all relevant Python functionalities corresponding to the equivalent Excel functionalities originally desired as an outcome of this course could be taught instead, so that students learn the same Data Science skills with Python instead of Excel (**Amir Shahmoradi**).
 - Instead, she recommends adding SQL to the curriculum, which is currently missing.
 - **Li Wang** is implementing her lecture notes in Jupyter notebooks and posts them on Blackboard.
 - **Li Wang** is teaching a combination of Data Science materials, from simple statistical and data manipulation tasks, to Python programming, to simple concepts in Machine Learning like K-means, Nearest-Neighbor, Random Forests, ... (the course material will be soon shared with

all). One advantage of such heterogenous course contents is that it can give the students an overview of the essential skills that the students would expect to learn throughout their degree program (**Amir Farbin**). In particular, students in the subsequent The It can One idea is to get the blackbox tools from Intro Data Science course to implement in Scientific computing.

- SQL currently seems to be missing in the list of courses to be developed/taught in the program. As an essential tool for manipulating large dataset, we may have to redesign the existing or add new course to include it in the curriculum (**Leili Shahriyari**).
- **The program website design:** The program's website has to be setup as soon as possible (**Dean Khaledi** , **Leili Shahriyari** , **Amir Farbin** , **Amir Shahmoradi**). This will be the discussion of the next meeting. Some initial guidance from the more senior college-level administrators or team-members will be certainly needed (via meetings) to ensure the website development follows the grand goals of the program (**Amir Farbin** , **Amir Shahmoradi**).

A good website development and design will be likely best achieved by hiring a dedicated web developer (**Amir Farbin** , **Amir Shahmoradi**).

- A shared repository will be setup for all the meetings' material (such as **Li Wang** 's course documents) that will be accessible to everyone in the team (**Amir Farbin**). Possibilities at the moment include Dropbox (**Leili Shahriyari**), UTBox (**Li Wang** , **Amir Farbin** , **Amir Shahmoradi** , **Leili Shahriyari**), or GitHub (**Amir Farbin**).
- **Possible Teaching Platforms:** Several teaching platforms were discussed between **Amir Farbin** , **Amir Shahmoradi** , and **Leili Shahriyari** over the summer:
 1. The Texas Advanced Computing Resources for Data Science (**Amir Shahmoradi**):
 2. Third-party clouds, such as OpenStack, Google, ... (**Leili Shahriyari**): Google provides it for free for teaching purposes.
 3. Amazon cloud, or alternatively, a dedicated UTA internal cloud server (**Amir Farbin**).
- Take advantage of UTA student help program (library) (**Leili Shahriyari**).
- The Intro Scientific Computing II should focus more on advanced high-performance programming languages (C/C++/Fortran) (**Daniel Welling** , **Amir Shahmoradi**).
- **Exams / Exercise / Homework:** How can plagiarism be minimized? (**Amir Farbin**) Still open-ended.