



Texas A&M Institute of Data Science

Thematic Data Science Labs Program

1. Aims of the Thematic Data Science Labs Program

The Texas A&M Institute of Data Science (TAMIDS) solicits proposals to establish Labs operating under the auspices of TAMIDS in emerging areas of Data Science (DS), here understood as including the foundations and applications of the associated fields of Machine Learning and Artificial Intelligence. The mission of each Lab will be to develop knowledge, resources, and community around an emerging thematic area of DS, encompassing research, education, and outreach. TAMIDS will support each Lab through a combination of seed funding for new research, effort from TAMIDS personnel, preferred access to existing TAMIDS programs, and organizational and logistical support.

Lab Activities: Each Lab will be initially focused around members who are Texas A&M investigators working coherently in the thematic area. The members of the Lab are expected to coordinate to: (i) conduct collaborative pilot research projects in the thematic area; (ii) develop training, tutorials, case studies and for-credit courses; (iii) establish data resources for use in research and education; (iv) organize workshops and seminar series; (v) recruit additional members through outreach to other investigators within their theme, in cognate disciplines, or in related application domains; (vi) involve students in Lab research and education activities; (vii) develop proposals for funding based on the work of the Lab. In short, the Labs program aims to build an ecosystem for Data Science around each thematic area that will make Texas A&M more competitive for external research funding.

Relation to X-Grants Program: The TAMIDS Thematic DS Labs program is independent of the X-Grants program and is focused exclusively on Data Science. The TAMIDS program explicitly requires community and resource development in addition to piloting new research. Priority in the 2021 TAMIDS program will be given to proposals which do not have substantial overlap in proposers or content with projects funded in the 2020/2021 cycle of the X-Grant; the TAMIDS deadline for full proposals occurs after X-Grants notification. The community and planning events required prior to TAMIDS proposal submission may enhance any future X-Grant submissions by their proposing teams.

2. Structure & Support of Thematic Data Science Labs

Lab Personnel: Each lab will be organized as follows:

- **Initial Members:** up to 8 PI-eligible Texas A&M members (see Q&A below for definition), with faculty members adloc'ed in at least 2 different colleges. The Lab is encouraged to recruit further members during its operation.
- **Leader (or Co-leaders):** initial member(s) who coordinate(s) Lab activities.
- **Postdoc:** TAMIDS-funded postdoc who collaborates with Lab members (see below).
- **TAMIDS Staff:** TAMIDS staff (in addition to the postdoc) may be involved in the Lab operation.

Lab Support: Each Lab will receive the following support from TAMIDS during each of its first two years of operation

- **Leadership:** \$10k support for leader, or split between co-leaders, which may be used to support salary.
- **Research:** \$20k towards each of two collaborative pilot research projects. Each project must involve at least two initial members from different departments and must pilot new research. The funds must be used for graduate student or postdoc support.
- **TAMIDS Postdoc:** a full-time postdoctoral researcher employed by TAMIDS to collaborate with initial Lab members. The Lab leader(s) may nominate the postdoc. The duties of the postdoc will be set by the Lab leader(s), in consultation with the TAMIDS director.
- **Outreach and Community:** \$5k for community building activities, such as a seminar series.
- **Programmatic Support:** additional support is available for Lab members through TAMIDS programs:



- **DS Course Development Program:** develop new or adapt existing DS courses; supported by research bursary payable to developer's a/c.
- **DS Tutorial Workshop Program:** develop and deliver 2-3 hour tutorials covering methods, tools, data, or hands-on case-studies; supported by a research bursary payable to developer's a/c.
- **DS Workshop Program:** Thematic workshop, half or full day, with TAMIDS supporting speaker expenses, and providing facilities if held in person.
- **DS Traineeship Program:** assistance on DS project development with assistance from TAMIDS for student engagement and mentorship; support available for a student worker.
- **Logistical Support:** Labs may use TAMIDS facilities for events and collaboration, and avail of TAMIDS administrative assistance as agreed with the TAMIDS Director.

Labs Community: during each year of funding, the Lab should extend its community through the following:

- **Workshop:** hold a thematic workshop supported by TAMIDS under its Workshop Grant Program. The workshop agenda might include: progress reports on pilot projects, other contributions from members or external speakers, planning discussions for the Labs program.
- **Seminar:** organize a seminar program of monthly talks relating to its theme.
- **Communications:** operate a segment of the TAMIDS website, a discussion group, or other social media to provide community focus for the Lab.

Sustaining Support, Review and Reporting:

- **Additional Funding:** the Lab leader(s) should work with TAMIDS to submit proposals to internal and external funding opportunities.
- **Sustaining TAMIDS Support:** The TAMIDS Lab postdoc should be included on all research proposals arising from the work that they perform for the Lab that are submitted while the Lab receives support and receive commensurate salary support if awarded.
- **Reporting:** The Lab leader(s) will deliver a Lab progress report each year at the thematic workshop event, and a final written report within 6 months of the end of the support period.
- **Review:** at the end of the first year TAMIDS and the Lab leader(s) will review and assess Lab outcomes and program effectiveness and make adjustments as needed.
- **Continuing Support:** TAMIDS may continue to support a Lab beyond the initial two-year period depending on outcomes from Labs activities.
- **Operating Agreement:** If awarded, the proposal, incorporating feedback from the TAMIDS review, will serve as the roadmap for operations and activities of the Lab.
- **Co-sponsorship:** Labs may be co-sponsored by other Texas A&M organizations that contribute significant resources to the Lab at its initiation or during its operation.

3. Procedure for Proposing a Thematic Data Science Lab

Establishment of Thematic Data Science Labs: The process to establish Labs occurs in three steps: Letter of Intent, Pre-submission Events, and Full Proposal.

- **Letter of Intent:** Prospective Lab leaders should submit a 2-page Lol in the format described below. Each Lol will be assessed for conformance to and expected fulfillment of program goals. Only Lol's selected by TAMIDS can proceed to full proposals.
- **Pre-submission Events:** at least three community events should be held prior to submission of the full proposal. These events should demonstrate: (i) coherence in the Lab theme of the research programs of the initial members; (ii) interest in the theme by investigators beyond the initial team; (iii) discussion or identification of educational and outreach activities that would be needed to promote the thematic area.
 - One of these events should be a thematic workshop of at least a half-day duration. The workshop can be held in person or online and must open to all PI-eligible members of Texas A&M. TAMIDS can provide support for the pre-submission workshop of selected Lol's under its Workshop Grant Program.
 - Other events such as RIG or planning meetings should have a combined duration of at least 4 hours. Participants, discussion & outcomes from the events should be documented for inclusion in the full proposal.
- **Full Proposal:** The Lab leader(s) should develop a full proposal in the format described below that sets out the specific Lab activities. Proposed Lab leaders are encouraged to discuss their proposal with the TAMIDS director. The proposal will be reviewed by a committee established by TAMIDS.

Letter of Intent Submission: The Lol should be submitted by email to datasciencejobs@tamu.edu by the

extended deadline of Friday, February 26, 2021, 12pm (noon) as a single pdf file comprising the elements below. Selection of Lols will be communicated by **March 15, 2021**.

- Title page comprising
 - Heading: “Letter of Intent for Thematic Data Science Labs”.
 - Lab Title: (100 characters max.)
 - Personnel: Name, affiliation, email address of the proposed leader(s) and other initial members, up to 8 PI-eligible in total with faculty members adloc’ed in at least 2 different colleges.
 - Project summary (200 words max.)
- One-page description (Arial 10pt, 1 inch margins all round) outlining the thematic area of Data Science and its potential for growth and impact on broad technical, societal, regional or national challenges; the current community of interest in Texas A&M around the theme; the anticipated pilot research program; expected education, outreach and resource development activities, potential funding opportunities.

Full Proposal Submission: Full proposals should be submitted by email to datasciencejobs@tamu.edu by the **deadline of July 2, 2021** as a single pdf file comprising the elements below. Awards will be announced by **August 16, 2021**.

- Title page comprising
 - Heading: “Proposal for Thematic Data Science Labs”.
 - Lab Title: (100 characters max.)
 - Personnel: Name, affiliation, email address of the proposed leader(s) and other initial members, up to 8 PI-eligible people in total. Faculty initial members cannot all have adloc to the same College.
 - Project summary (200 words max.)
- Separate page listing further possible Lab members based on e.g. research area interest, participation in pre-submission activities: Name, affiliation, email address.
- Project description comprising no more than eight pages of text, 10-point Arial font, 1 inch margin all around. Two additional pages of figures may be included. The proposal must include the following sections:
 1. Labs Thematic Area: What is the thematic area within DS that the Lab addresses? What is the compelling case for this theme to be developed, in terms of research opportunities and impact on broad technical, societal, regional or national challenges? Include references that document the emergence or potential for growth in the proposed thematic area.
 2. Report on Pre-submission Activities: Document the pre-submission events: dates, event type, who attended (e.g. speakers, # attendees, other key contributors), what was discussed, outcomes & conclusions from the discussions that motivate the Labs proposal.
 3. Team Coherence: Describe the coherence of the initial team for the Lab theme, e.g. through existing research and/or expected to emerge through pilot projects, use of common or similar resources such as: systems, data or methods, and; linkages through common interest in broad technical, societal, regional or national challenges.
 4. Research Program: What is the overall goal of the proposed Data Science pilot research program? How are results and capabilities generated by the pilot projects within the two-year funding period expected to position the Lab for future research?
 5. Pilot Projects: Two pilot projects will be funded. The proposal may list addition unfunded pilot projects, which may avail of postdoc effort. Funded projects involving multiple members should designate one member as leader, who will be responsible for allocation of project funds. For each pilot project describe:
 - a. Pilot project title
 - b. Problem description, novelty, and fit with thematic area
 - c. Project approach, outcomes, results and products expected during year 1.
 - d. Any data resources involved
 - e. Initial Lab members involved and their preparation for project (these members may not all come from the same department).Outline potential continuations and/or additional pilot projects for year 2.
 6. Education and Training: Describe courses, training, tutorial and any other educational resources that will be developed in the Lab, which initial member(s) will be involved, and TAMIDS programs that will be used to support this activity. Who is the target audience? What gaps in current programs will these resources fill?
 7. Community Growth and Impact: List principal organizers of the seminar program and annual workshop and any innovations for these events. Describe other proposed community building actions. How will these actions

encourage participation in the Lab from students and underrepresented groups? What new interdisciplinary interactions, or interactions between foundations, applications and/or practice will be stimulated by the Lab? What data, computational, or other resources will be developed, which Lab members will develop them, and how will they help the community? How will new members be recruited to the Lab and what benefits will accrue to them through participation? Describe any anticipated broader impacts that the Lab will have beyond its membership in Texas A&M, and outside the university. What new external relationships will the Lab develop?

8. Potential for External Funding: Explain how the Lab program can make TAMU more competitive for external funding. What Texas A&M and external funding sources does the Lab team plan to compete for, including names of relevant programs.
 9. TAMIDS Postdoc: Describe the role of the TAMIDS postdoc in the Lab including any activities in pilot projects, education and training. The postdoc may also work in non-pilot projects that are led by initial members and consistent with the Lab theme. If this is planned, the proposal should give project title, initial member lead(s), and a brief project description. The postdoc should not be tied exclusively to a single project or Lab member and should be able to act as an agent for collaboration and cross-fertilization within the Lab. The proposal should specify postdoc skills and experience that are required or preferred. The proposal should include a resume of any person the Lab leader(s) wish to nominate for this role. Any role for additional TAMIDS staff should also be described here.
 10. Collaborating Organizations: proposers are encouraged to secure Texas A&M organizational co-sponsors for the Labs that will contribute significant additional resources. For any collaborating organizations, the proposal should list of participating personnel, their roles, and resources and/or capabilities to be provided.
 11. (Optional) Other Information. Any other information the proposers wish to be considered.
- References cited in the proposal (No page limit).
 - Collaboration Letters: Letters of intent to collaborate with the Lab should be included from:
 - a. Any organizations co-sponsoring the Lab, confirming personnel, their roles, and resources and/or capabilities to be provided.
 - b. Any non-sponsoring organizations that will be involved in Lab activities
 - Two-page resume for initial Lab members (following the NSF biosketch format) and any proposed candidate for the TAMIDS postdoc.

4. Questions and Answers

- What is the total financial support for a Thematic Data Science Lab?
 - *Direct funding to researchers and leaders, TAMIDS postdoc & personnel, and TAMIDS program support totals about \$300k over 2 years.*
- How many thematic Labs will be funded?
 - *TAMIDS expects to fund a minimum of new 2 Labs in the 2021 program.*
- How many Letters of Intent will be selected for a full proposal?
 - *TAMIDS expects that around 8 Lol's will be selected.*
- Does the Lab title and initial membership need to be identical for the full proposal and the Lol?
 - *No. But substantial departures of the full proposal from the approved Lol may make the proposal less competitive if it diverges from program goals.*
- Which Texas A&M units may Lab members belong to?
 - *This program is open to all PI-eligible faculty, researchers, and staff at Texas A&M University (including Galveston and Qatar campuses), TEES, TEEX, AgriLife Research, AgriLife Extension, and TTI.*
- Why is there a limit on the number of initial Lab members?
 - *The Lab should have focused interlocking goals for research, education, resource and community development, with specified members taking on these roles. The limit aims to promote tight coordination between initial Lab members.*
- Are any Thematic Data Science Labs currently operating?
 - *The Operational Data Science Lab was established jointly with the Texas A&M Center for Geospatial Science, Applications, and Technology (GEOSAT) in 2020, and the Scientific Machine Learning Lab was established in 2021 as a pilot for this program; see <https://tamids.tamu.edu> for details.*
- Who is the TAMIDS contact for the program?
 - *Jennifer South, TAMIDS, jsouth@tamu.edu*