

## **CHARTER RENEWAL APPLICATION SIAM Activity Group on Data Mining and Analytics (SIAG/DMA) - Proposed Rescope to SIAM Activity Group on Data Science (SIAG/DATA)**

This CHARTER RENEWAL APPLICATION applies to the SIAM Activity Group on Data Mining and Analytics (SIAG/DMA). The SIAG/DMA was originally formed under the aegis of SIAM on July 17, 2011 by the SIAM Council and by the SIAM Board of Trustees. The SIAG has been renewed by the council and board four times between 2011 and 2019. The SIAG had 1,289 members as of December 31, 2019; of these, 947 were students.

The main goal of this application by the current officers of the SIAM Activity Group on Data Mining and Analytics (SIAG/DMA) is to propose to rescope SIAG/DMA to serve the broader data science community and change its name into “ **SIAM Activity Group on Data Science (SIAG/DATA)**, ” along with rescoping the aims and goals of the SIAG to fully encompass the emerging field of data science, opening up from the more specialized “data mining” subarea of the field of data science that was the focus of SIAG/DMA.

This proposal for broadening, rescoping and renaming the SIAG is being submitted to the SIAM council and board after extensive discussions and consultations between 1) current and former officers of SIAG/DMA, 2) the steering committee of the annual SIAM International Conference on Data Mining (SDM), which is the flagship activity of SIAG/DMA, 3) the inaugural co-chairs of the new SIAM Conference on Mathematics of Data Science (MDS), 4) editors of the new SIAM Journal on Mathematics of Data Science (SIMODS), 5) recent officers of other SIAGs that have an interest in data science (SIAG/IS, SIAG/OPT, SIAG/CSE, SIAG/LA, SIAG/SC), and 6) SIAM leadership.

The context for the current proposal is as follows.

SIAG/DMA has been a thriving activity group with a strong membership record and a highly selective annual conference of high academic standing (the SIAM International Conference on Data Mining, SDM), focusing on the specialized area of data mining. However, in recent years, the broader area of “data science” has taken off tremendously in terms of reach and scope, with a strong and growing impact in basic and applied research and many new groundbreaking applications across society. Mathematics, statistics and computer science lie at the foundation of data science, and as such, many in the community realized SIAM needs a dedicated SIAG home for the quickly growing field of data science.

As such, the newly scoped SIAG will complement other recent SIAM initiatives in the data science field, including the new Mathematics of Data Science conference (which was originally scheduled to debut in May 2020, with accepted 158 minisymposia, 152 contributed talks, and 38 poster submissions, but had to be postponed due to the covid-19 pandemic), and the new SIAM Journal on Mathematics of Data Science (SIMODS), which launched in 2019.

Rather than forming a new SIAG/DATA in addition to the existing SIAG/DMA, there is a consensus among all consulted stakeholders and community members that the natural and best way forward is the broadening, renaming and rescoping of SIAG/DMA into SIAG/DATA. The current SIAG/DMA officers support and lead the initiative for renaming and rescoping the SIAG. The SDM conference community (through the steering committee) also expressed a desire to remain included in the broader data science SIAG (because data mining is a part of data science), rather than retaining a dedicated SIAG/DMA (which would likely end up substantially depleted in terms of membership). Similarly, the consulted stakeholders from the MDS conference, the SIMODS journal, and recent officers of the other SIAGs, unanimously supported forming the SIAG/DATA as a continuation and broadening of SIAG/DMA.

There was substantial debate within the group we consulted regarding the choice of name for the broadened SIAG, between the two options “SIAG on Data Science” and “SIAG on Mathematics of Data Science”. The main advantage of the second option is a clearer focus on mathematical aspects of data science (which is useful to set SIAM’s data science group apart within a data science field that is now so broad that it encompasses many areas of science and technology). The main advantage of the first name is inclusivity and perhaps to better reflect the broad and interdisciplinary scope of SIAM’s membership, that includes about 50% of professional mathematicians, but also substantial fractions of computer scientists, engineers, physicists, etc. In the end, a majority of the participants in the discussion felt the name “SIAM Activity Group on Data Science” was preferable, also since the broader name is consistent with the desire to retain the SIAM Data Mining conference as a key activity of the SIAG (the audience for this conference has a large computer science component), and realizing that the word “SIAM” in the title already indicates sufficiently that this is an activity group within a mathematics-focused professional society.

It was also agreed among all stakeholders mentioned above that the flagship activity of the new SIAG/DATA will be the bi-annual SIAM Conference on Mathematics of Data Science (MDS) (from 2022). The program director and chairperson of SIAG/DATA will be responsible for ensuring MDS’s continuation. The program director or designate will usually be a co-chair of the conference, and additional co-chairs will be selected by the chair, the program director and the other officers of the SIAG. The SIAG/DATA officers will actively pursue broad representation in the set of MDS co-chairs and the MDS organizing committee from all SIAGs that have an interest in the broad area of data science. Similarly, the nomination committees for electing SIAG/DATA officers will also pursue such broad representation.

In addition to its new flagship bi-annual Mathematics of Data Science conference, SIAG/DATA will also continue to sponsor the annual SIAM International Conference on Data Mining (SDM) and work with the Steering Committee to support its organization. Continuing current practice, SDM will remain under the main guidance of the SDM Steering Committee. To ensure information flow between the SIAG officers and the SDM

conference, the SIAG chair or designate will serve as an observing ex-officio member on the steering committee of the SIAM International Conference on Data Mining (SDM).

The SDM steering committee chair will participate in the new SIAG officer orientation meeting at the beginning of the term of the new officers, to inform relevant parties about the status and needs of the SDM conference. The steering committee chair will provide input about the SDM conference for charter renewal and the SIAG chair will provide input for the SDM business meeting about the activity group and the MDS conference.

Since this rescoping of SIAG/DMA into SIAG/DATA can in many ways be viewed as an activity group making a new start, we would like to propose (after discussion with Vice President-at-Large Carol Woodward and Executive Director Jim Crowley) to appoint a slate of officers for the first year of operation of the new SIAG (from 1/1/2021 to 12/31/2021), similar to the procedure for a new SIAG. For one, this avoids cumbersome complications with holding a vote for the first slate of officers, since it would not be easy to have new members join the SIAG following the rescoping and renaming and having them participate in an election, before the SIAG's approval and reformation under its new name. While an appointed slate of officers would usually stay on for two years, we believe a one-year term for appointed officers may in our case be the best choice, due to alignment with the planned conference schedule for the flagship MDS conference of the SIAG (MDS-22, MDS-24, MDS-26, etc.). Specifically, if the appointed officers hold an election in 2021, the newly elected program director for 1/1/2022 can become a co-chair for the next MDS conference, planned for May 2024, and gain the opportunity to observe how MDS-22 unfolds in May 2022. The program officer elected for 1/1/2022, together with the other officers, would also oversee site selection for the 2026 MDS conference, since conference locations for large conferences need to be secured three or more years in advance. Similarly, the program director of the appointed officer slate for 2021 will be a co-chair for the 2022 MDS conference, and will, together with the other officers, be responsible for site selection of MDS-24, and the program director elected for 1/1/2024 will be a co-chair for MDS-26, and will select, with the other officers, the site for MDS-28.

After enacting this renaming and rescoping, SIAG/DATA proposes to operate according to the following answers to the usual renewal questions for SIAGs:

The aim of the SIAM Activity Group on Data Science is to advance the mathematics of data science, to highlight the importance and benefits of data science, and to identify and explore the connections between data science and other applied sciences. The SIAG will focus on the mathematical, statistical and computer science foundations of data science, and on applications of data science to other fields of science and across technology and society. Within the framework of SIAM, the SIAG will conduct activities that implement its purposes.

The SIAG on Data Science organizes activities in Data Science. The SIAG is expected to:

1. Organize minisymposia at the SIAM Annual Meeting.

2. At least once every seven years, either organize a track of at least six mini-symposia at the SIAM Annual Meeting or have an activity group meeting held jointly with the annual meeting. The VP for Programs and the VP at Large will coordinate the scheduling with the SIAG chair.
3. Organize a biennial conference on the Mathematics of Data Science (MDS). The program director and chairperson of the SIAG are responsible for ensuring the continuity of MDS. The program director or designate will usually be a co-chair of the conference, and additional co-chairs will be selected by the chair, the program director and the other officers of the SIAG. The co-chairs and organizing committee must be approved by the VP for Programs at least 16 months before the conference.
4. SIAG/DATA also sponsors the annual SIAM International Conference on Data Mining (SDM) and works with the Steering Committee to support its organization.

The SIAG will complement SIAM's activities and support its proposed functions. The answers to the questions below indicate how this will be accomplished and what the officers propose as the future directions for the SIAG.

I. How is the field covered by the activity group doing? Is it growing? Is the focus shifting? What have been the significant advances over the past two years?

As discussed above:

SIAG/DMA has been a thriving activity group with a strong membership record and a highly selective annual conference (SDM) of high academic standing, focusing on the specialized area of data mining.

However, in recent years, the broader area of “data science” has taken off tremendously in terms of reach and scope, with a strong and growing impact in basic and applied research and many new groundbreaking applications across society. Mathematics, statistics and computer science lie at the core of data science, and as such, many in the society realized that SIAM needs a dedicated SIAG home for the quickly growing field of data science.

For this reason, we and the stakeholders we consulted propose to broaden and rename the SIAM Activity Group on Data Mining and Analytics (SIAG/DMA) into the SIAM Activity Group on Data Science (SIAG/DATA).

2. How is the activity group doing? Is it remaining vibrant? Is the size of the SIAG stable or increasing? How is the SIAG keeping up with the changes in the field? How are the broader interests of SIAM reflected in the activities of the SIAG?

The membership of SIAG/DMA has been growing, and we expect it will continue to grow further after the broadening and renaming of SIAG/DMA into SIAG/DATA.

As proven by the success of the new SIAM Conference on Mathematics of Data Science (MDS) and the new SIAM Journal on Mathematics of Data Science (SIMODS), there is tremendous potential growth for SIAM in the data science area.

3. Please list conferences/workshops the activity group has sponsored or co-sponsored over the past two years and give a brief (one sentence or phrase) indication of the success or problems with each.

The new SIAM MDS conference and the existing SDM conference were scheduled to be co-held May 5 - 9, 2020 in Cincinnati, Ohio, US. SDM was on track to maintain its size and reach compared to previous years (with 312 submissions and 75 accepted high-quality papers now published in the conference proceedings), and MDS was on track to become a large and high-quality conference even at its first instalment, with 158 minisymposia, 152 contributed talks, and 38 poster submissions and an outstanding list of plenary speakers. We anticipated co-locating these conferences to enhance substantial new interactions and create synergy between the specialized data mining community and the broader SIAM data science community. Unfortunately, the co-located conferences had to be postponed or cancelled due to the covid-19 pandemic. Depending on the covid-19 situation, the next instalment of SDM may be held in 2021, and the next instalment of MDS is planned for May 2022 (likely again in a co-located setting, with SDM-22).

Since the first MDS conference had to be canceled, the co-chairs are currently exploring holding a smaller virtual event later in 2020 with a small number of high-profile invited talks and a few sessions where graduate students can present their results (we prioritized a venue for graduate students since their graduation timelines are often not extended as a result of the pandemic and it is important for them to get opportunities to present their results).

4. Please indicate the number of minisymposia directly organized by the activity group at the last two SIAM Annual Meetings. When did the SIAG last organize a track of minisymposia at an annual meeting?

SIAG/DMA has not directly organized minisymposia at the two most recent SIAM annual meetings or ICIAM. The broadening to SIAG/DATA will be an excellent opportunity to seek a special track at the next SIAM Annual Meeting, and the broadened SIAG plans to actively pursue this. Since data science is now such an active research topic and SIAM has many members in its various subdisciplines who are actively engaging in exciting new research in the data science area, we expect there will be substantial interest from members of the renewed SIAG to present their work at the SIAM annual meeting, where we believe it will find a very receptive audience, considering the current broad interest across the society in data science topics and results.

5. Please indicate other activities sponsored by the activity group, to include newsletters, prizes and web sites. Have each of these been active and successful?

SIAG/DMA has maintained a very active mailing list for its members, with job postings in academia and industry, conference and workshop announcements, graduate student opportunities, etc.

The SIAM International Conference on Data Mining features the annual IBM Early Career Data Mining Research Award, a plenary Doctoral Student Forum, and a Best Research Paper Award.

6. What activities are planned and proposed for the next period of the charter? Please describe scheduled and suggested future activities in detail.

As discussed above, the flagship conference of the rescoped and renamed SIAG/DATA will be the bi-annual SIAM Conference on Mathematics of Data Science (MDS), scheduled for May 2022. The SIAG will also continue to sponsor the annual SIAM International Conference on Data Mining (SDM), planned for 2021. The 2022 MDS and SDM conferences will likely be co-located, to enhance the cross-fertilization between the data mining community and the broader SIAM data science community.

The renewed SIAG/DATA will actively pursue broadening of its offerings to its membership, which we believe will quickly grow very substantially. The scope and topics of the existing email list will broaden, and we expect it will become an important resource

for the data science community across SIAM in terms of job postings in academia and industry, conference and workshop announcements, graduate student opportunities, etc.

SIAG/DATA will also embark on an initiative to create SIAG prizes for research excellence in the data science area. We will start with a SIAG/DATA Early Career Prize, to be established in early 2021, targeting first awarding at the 2022 MDS conference. In our preliminary planning, this will be followed by a SIAG/DATA best publication award for MDS-24, and a SIAG/DATA career prize for MDS-26.

SIAG/DATA will also consider starting a regular newsletter for the community (possibly modeled after SIAG/OPT Views-and-News).

7. How can SIAM help the activity group achieve its goals?

We believe there are tremendous opportunities for SIAM in establishing an activity group that can be the overall home for data science in SIAM. We appreciate the support of the council and board in considering our request to rename and rescope SIAM/DMA into SIAG/DATA to start to fulfil that role.

8. How can the activity group help SIAM in its general role of promoting applied mathematics and computational science?

Data science has quickly become one of the great quantitative endeavors of our time, with tremendous innovations in all areas of science, technology and society-at-large that are based on foundations in mathematics, statistics and computer science. SIAM has made important strides in the data science field with the new SIMODS journal and the new MDS conference. Complementary to those recent SIAM initiatives, establishing a dedicated activity group home for data science in SIAM, in the form the rescoped SIAG/DATA may fulfil, is, in fact, overdue, and will be a powerful vehicle for SIAM in promoting the crucial role applied mathematics and computational science play in data science advances across society.

SIAG/DMA requests that the SIAM Council and Board of Trustees establish/renew the charter of the rescoped SIAG/DATA for a one-year operating period beginning January 1, 2021. (Note that this is shorter than the usual two-year cycle, but a one-year period is consistent with the duration of appointment for the first slate of officers of the rescoped SIAG/DATA.)

Ali Pinar, Chair of the SIAG on Data Mining and Analytics  
Hans De Sterck, Vice-Chair of the SIAG on Data Mining and Analytics  
May 22, 2020