ADDENDUM

Global Leadership in Data-intensive Discovery

Leadership in data-intensive discovery is “the rising tide that lifts all boats” — the single investment, we believe, that will have the greatest impact on the UW’s future competitiveness. We also deeply believe this leadership is within our grasp. In this addendum to our proposal, we focus on how and why WRF’s support will be transformative, the relationship of this support to other commitments, our targets and metrics, and our revised budget.

Elements of the budget

Our proposal requests funding for chairs, professorships, and startup packages to assist in recruiting game-changing faculty; for postdoctoral fellowships to generate rapid impact; and for remodeling and furnishing the 6th floor of the Physics/Astronomy Building as a campus Data Science Studio to foster cross-disciplinary collaboration and entrepreneurship.

Chairs, professorships, and startup funds: We referred in our proposal to a virtuous cycle in which advances in data science methodologies transform the process of discovery and drive new discoveries in a broad range of fields of great importance to the UW, WRF and our partners, while at the same time the needs of discovery in these fields stimulate the creation of new data science methodologies. With this goal in mind, the provost has invested in a visionary program of “social engineering.” Under the Provost’s Initiative in Data-Intensive Discovery, she has allocated to the eScience Institute $300,000 in permanent annual funds, and she has committed an additional $150,000 in permanent annual funds, in the form of six “half faculty positions.” When a unit seeks to hire an individual with strong credentials on both sides of this virtuous cycle, we can provide 50% support for this individual with the stipulation that 50% of his or her effort is devoted to “broad impact” activities in data science — for example, in the education realm, teaching courses in our new NSF IGERT interdisciplinary graduate program in data science. The challenge is persuading these extraordinary individuals, for whom there is tremendous competition, to choose the UW. Two elements are required, for neither of which the UW has resources:

- **Chairs and professorships.** Jeff Bezos’ creation of the Amazon Professorships in Machine Learning was central to our success in recruiting Emily Fox and Carlos Guestrin two years ago. Chairs and professorships carry both prestige and funding. The funding, although relatively modest on the scale of federal research grants, is tremendously valuable because it can serve as a “seed fund” for exploratory studies in new directions. We propose to establish three WRF Data Science Chairs and three WRF Data Science Professorships, each to last five years.

- **Startup funding.** New faculty must establish labs. New labs require funds for remodeling, equipment, and bootstrapping staff and students as grant funding ramps up. In Computer Science & Engineering, one of the least expensive fields because lab requirements are comparatively modest, we routinely spend $400,000 per recruit on startup funding. (WRF has provided crucial assistance for this in the past — for example, in helping Shwetak Patel get established.) There is no adequate internal source for these funds. We have proposed six startup packages.
**Postdoctoral fellowships:** In most scientific fields it is common for an individual to spend time after the Ph.D. as a postdoctoral fellow. This is an important stage in the transition from being “someone’s student” to becoming an independent contributor and leader in academia or industry. For the established faculty member who is a postdoc’s mentor, it is a major impact multiplier for three reasons:

- Postdocs hit the ground running, as opposed to graduate students, who are learning how to do research. Postdocs already know how to do research and are on a path to independence.
- Postdocs move on to permanent positions after several years, “spreading the gospel” rapidly. This is particularly critical in data science, where we are trying to change ways of discovery.
- Postdocs — particularly if dual-mentored, as we intend — can serve as a critical technology transfer bridge, either between scientific fields, or between university and industry.

Our funding from the Moore and Sloan foundations is largely for professional staff — research scientists and data scientists who will work with faculty and students to help them extract knowledge from their data. Our funding from the NSF IGERT is for graduate students. Neither source is a funding mechanism for postdocs. Thus, we propose to establish three cohorts of WRF Data Science Postdoctoral Fellows, with each fellowship lasting three years. Each postdoc will have two mentors — one from a data science methodology field and one from a data science application field in academia or industry — providing a vibrant network for collaboration. We would suggest that each project supported by WRF name their postdoctoral fellows similarly, and that we have technical, social, and mentoring events involving the postdocs from all WRF projects, enhancing interdisciplinary interactions and creating a prestigious “branded” postdoctoral program. Our project team would be pleased to take the lead in this.

**Data Science Studio:** In the days of punched card computing, we would talk to others while hanging around the computer center waiting for our listing to emerge from the bowels of the machine room. And we would discover that, although we were in different fields, we shared similar solutions. When computing retreated to the desktop and then to the cloud, the “water cooler effect” disappeared. We will recreate it by establishing a physical place where faculty and students can interact with one another and with professional research scientists and data scientists. Data science is the interdisciplinary bridge for the foreseeable future. The UW has allocated the top floor of the centrally located Physics/Astronomy Building — currently the Physics/Astronomy Library — for this purpose, and the Moore and Sloan foundations have provided funding for the staff and programs that will anchor this space. We have carried out a preliminary design for a remodel, but we have no funds to actually carry out the remodel. We propose to utilize WRF funding for this purpose, and to name this campus-wide interaction space the **WRF Data Science Studio**.

**Institutional commitment**

In choosing the UW as a partner in the Data Science Environments initiative, the Moore and Sloan foundations were keenly focused on the depth of commitment — both to data science and to institutional change — exhibited by central, school/college, and departmental leadership. Our campus-wide data science event on February 7 was headlined by the president, the provost, and the vice provost for research; more than 130 individuals and organizations contributed research posters to an extraordinary networking session. Core funding to establish the eScience Institute was obtained from the Legislature (currently $500,000 per year plus one full faculty position at $150,000 per year). Incubation space was allocated by Computer Science & Engineering. Six new faculty positions are now available through the Provost’s Initiative in Data-Intensive Discovery, with a permanent funding level of $450,000 per year from the provost and an additional $450,000 per year (minimum) from the departments that hire these individuals. In addition, the departments will provide the research and office space for these hires plus any startup funds that may be required beyond those requested from WRF. The Physics/Astronomy Library space has been allocated as a Data Science Studio, a major
commitment of prime space in what is a terribly space-constrained environment. The Office of Research gave our team the highly sought after UW “chit” to compete for the NSF IGERT. Funding of this WRF initiative will ensure that these commitments are highly leveraged.

Progress targets
We believe we can achieve the admittedly aggressive targets of our proposed budget. Specifically:

- **Faculty.** Because the Provost’s Initiative is in place, various units are currently interviewing appropriate candidates. The challenge will be successfully luring the ones to whom we choose to extend offers. At the seniority level appropriate for professorships, two candidates being interviewed jointly by Statistics and Computer Science & Engineering and one being interviewed by Computer Science & Engineering are precisely in the “sweet spot.” At the seniority level appropriate for chairs, one candidate being interviewed by Astronomy would be an outstanding fit. There will not be a perfect match of people to positions in any single year, but we are not at a standing start. Importantly, the Provost’s Initiative encourages UW units to make the commitment to hiring “Pi-shaped” faculty members (strong in both methodologies and applications). The WRF-funded chairs, professorships, and startup packages will be critical in persuading these individuals to choose the UW.

- **Postdocs.** Here again the wheels are already turning. The UW, Berkeley, and NYU have small numbers of postdocs funded under the Moore/Sloan initiative. We have advertised broadly and have received applications from a significant number of outstanding candidates — far more than the small number of positions we have available. To ensure we are able to assimilate and properly mentor new people during our startup phase, we have reduced the proposed size of the first cohort of WRF Data Science Postdoctoral Fellows from five in the original budget to three in the new budget.

- **Space.** Design of the remodel is underway and is a collaborative process among the eScience Institute, the UW Libraries, Physics, Astronomy, and the College of Arts & Sciences. The Physics/Astronomy Library collection will be removed during spring quarter, and remodeling can commence in June if funds are available. Our latest estimate is that remodel expenses will be closer to $900,000 than the $750,000 we originally budgeted; in reducing the number of postdocs in the first cohort from five to three, we have shifted the first-year funding into remodeling and decreased the second and third year budgets.

From a project monitoring point of view, WRF’s investment will coincide almost precisely with the Moore/Sloan, NSF IGERT, Provost’s Initiative, and UW Libraries investments. Thus, the effort to ensure that the UW is a leader in data-intensive science will have plenty of eyes focused on it. Following are the metrics we will apply; the first three are quantitative and can be directly tied to the release of funds:

- Are appropriate faculty being recruited at the planned pace?
- Are appropriate postdocs being recruited at the planned pace?
- Is the remodeling carried out efficiently and effectively?
- Is the WRF Data Science Studio a vibrant cross-campus interaction space?
- Are the various programs planned for that space effective? (Participation in “boot camps,” numbers of successful software “incubation projects,” etc.)
- More broadly, are we successfully creating a vibrant campus-wide community in data science?
- Can we trace to this initiative a progression of new approaches to discovery, and new discoveries?
- Are the new educational offerings planned as part of the NSF IGERT on track?
- Is tech transfer and entrepreneurship thriving?
- Is the UW attracting national and international recognition for leadership in data science?
Complementary funding sources
We will discuss complementary funding sources and purposes during the site visit. We hope, though, that the preceding discussion has clearly indicated that WRF’s support will meet specific, critically important needs that are not met by our current funding sources, and that WRF’s support will be transformative in terms of the impact we can achieve. From the establishment of the eScience Institute in 2008 until our success in the intense IGERT and Moore/Sloan competitions in the past six months, we have been a scrappy startup with a small but extraordinary team that has accomplished an amazing amount on a relative shoestring — a core budget of less than $400,000 per year. Our team’s success in positioning the UW for leadership in data-intensive discovery — combined with the impact this leadership will have — is precisely what makes us a strong investment target for WRF.

Revised budget
The budget below reflects the following modifications from our original version:

- Fund each faculty startup package over two years vs. a single year. (It is reasonable to expect two years to be required to expend these funds.)
- Reduce the number of postdocs in the first-year cohort from five to three, putting the additional first-year funds into remodeling and reducing the second- and third-year budgets appropriately.
- Extend the five-year funding of the second- and third-year chairs and professorships into sixth and seventh years vs. accumulating the necessary funds during the five years of the project.

<table>
<thead>
<tr>
<th>3 5-year chairs ($100,000/yr)</th>
<th>Year 1</th>
<th>Year 2</th>
<th>Year 3</th>
<th>Year 4</th>
<th>Year 5</th>
<th>Year 6</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 new</td>
<td>$100,000</td>
<td>$200,000</td>
<td>$300,000</td>
<td>$300,000</td>
<td>$300,000</td>
<td>$300,000</td>
<td>$1,500,000</td>
</tr>
<tr>
<td>3 5-year professorships ($50,000/yr)</td>
<td>1 new</td>
<td>1 new</td>
<td>1 new</td>
<td>Year 1</td>
<td>Year 2</td>
<td>Year 3</td>
<td>Year 4</td>
</tr>
<tr>
<td>1 new</td>
<td>$50,000</td>
<td>$100,000</td>
<td>$150,000</td>
<td>$150,000</td>
<td>$150,000</td>
<td>$150,000</td>
<td>$750,000</td>
</tr>
<tr>
<td>6 startup packages ($500,000 over 2 yrs)</td>
<td>2 new</td>
<td>2 new</td>
<td>2 new</td>
<td>Year 1</td>
<td>Year 2</td>
<td>Year 3</td>
<td>Year 4</td>
</tr>
<tr>
<td>$500,000</td>
<td>$1,000,000</td>
<td>$1,000,000</td>
<td>$500,000</td>
<td>$800,000</td>
<td>$400,000</td>
<td>$3,120,000</td>
<td></td>
</tr>
<tr>
<td>13 3-year postdoctoral fellows ($80,000/yr)</td>
<td>3 new</td>
<td>5 new</td>
<td>5 new</td>
<td>Year 1</td>
<td>Year 2</td>
<td>Year 3</td>
<td>Year 4</td>
</tr>
<tr>
<td>$240,000</td>
<td>$640,000</td>
<td>$1,040,000</td>
<td>$800,000</td>
<td>$400,000</td>
<td>$3,120,000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Remodel and furnish Data Science Studio</td>
<td>$900,000</td>
<td>$900,000</td>
<td>$900,000</td>
<td>$900,000</td>
<td>$900,000</td>
<td>$900,000</td>
<td>$900,000</td>
</tr>
<tr>
<td>Annual total</td>
<td>$1,790,000</td>
<td>$1,940,000</td>
<td>$2,490,000</td>
<td>$1,750,000</td>
<td>$850,000</td>
<td>$450,000</td>
<td>$9,270,000</td>
</tr>
</tbody>
</table>