**Post-Mortem**

**SICSS-Chile 2023**

## Introduction

SICSS-Chile 2023 was held at the Pontificia Universidad Católica, Campus San Joaquín, with the support of the Millennium Institute for Foundational Research on Data [(IMFD)](https://imfd.cl/), the National Center For Artificial Intelligence [(CENIA)](https://cenia.cl/), the Computer Science Department of the University of Chile, and the Institute of Sociology of the Pontifical Catholic University of Chile (ISUC). The event was organized by five organizers (Andrés Abeliuk, Hernán Sarmiento, Pedro Seguel, Pablo Beytía, and Juan Reutter). With four Teaching Assistants (Sebastían Ferrada, Jorge Ortiz, Gabriel Iturra, and Cinthia Sánchez). We hold all sessions in-person with 20 participants for two weeks.

## Applications and Outreach

We received applications from 62 eager candidates, from which 20 were selected. The selection process entailed each member of our five-person organizing team nominating 10 candidates based on the strength of their applications. After creating this preliminary list, we engaged in a comprehensive discussion to finalize our selections, ensuring we maintained gender and disciplinary diversity.

Upon finalizing our selection, we extended invitations to the chosen candidates, seeking their confirmation to participate in the nine-day SICSS event. If any selected individual was unable to commit, we proceeded to invite the next deserving candidate from our list.

Feedback from participants revealed diverse channels through which they discovered the event. Notably, many cited LinkedIn posts by organizers and associated organizations. Others were informed by their supervisors, discovered the event on the SICSS website's location pages, or heard about it through individuals affiliated with our partner organizations.

## Onboarding

To ensure participants were well-prepared, those who confirmed their attendance were provided with several resource recommendations:

1. The book “Bit by Bit: Social Research in the Digital Age” by Matt Salganik was listed as a primary read.
2. We recommended the standard SICSS boot camp videos.
3. A special SICSS-Chile was announced, which would be made available a week before the Institute's commencement.
4. In alignment with our decision to primarily use Python and Collab for our software tools, we suggested participants undertake the course '[Introduction to Computational Social Science with Python](https://github.com/gesis-css-python/materials)'. This course, crafted by Dr. Milena Tsvetkova and Dr. Patrick Gildersleve from GESIS, offers a comprehensive collection of materials, exercises, and references specific to Computational Social Science in Python.
5. We also highlighted online resources from our research partners CENIA and IMFD, which are predominantly in Spanish:
   1. IMFD Youtube Channel: This platform showcases numerous talks, keynotes, and tutorials on both practical and theoretical aspects of data management.
   2. OpenCenia: A resource-rich website offering courses, datasets, and AI models.
6. Three weeks prior to the event, participants were added to a dedicated Slack group, where they were encouraged to introduce themselves and familiarize themselves with the team.

Despite disseminating this information through e-mails and the Institute's website, participants did not explore the resources extensively. A potential reason might be the evident gap in coding skills among attendees; this is inferred from the observation that the initial boot camp workshops had higher attendance compared to subsequent sessions. The intensive coding component may have been daunting for some attendees.

## First Week

The first week we spotlighted guest speakers, foundational lectures, and hands-on code workshops, complemented by networking activities. The week culminated on Friday with idea brainstorming and project group formation, spanning topics from data ethics to visualization and natural language processing.

The detailed activities of the first week are described as follows:

* **Monday, July 17, 2023**. The day commenced with activities designed for participants to get acquainted. Attendees prepared a 2-minute slideshow presentation. Andrés Abeliuk provided an overview of Computational Social Science, leveraging the SICSS curriculum and beyond. Claudia López highlighted Data and Ethics in her keynote, followed by an ethics case discussion adapted for the Spanish language and Chilean context by Pedro Seguel.
* **Tuesday, July 18, 2023**. The focus was on textual sources and NLP tools for social research. The day started with Sebastián Ferrada's lecture and workshop on Knowledge graphs and Querying Wikidata. This was followed by Pedro Seguel's introduction to 'Text as Data/NLP' and Jorge Ortiz's workshop on basic text analysis techniques, covering pre-processing and bag-of-words analysis.
* **Wednesday, July 19, 2023**. The event's third day was dedicated to exploring Text Analysis and Natural Language Processing. The day began with a workshop on "Text Analysis II", expertly led by Gabriel Iturra. This workshop introduced word embeddings models for describing and classifying data. Subsequently, attendees were then captivated by a keynote on "Large Language Models" delivered by Felipe Bravo-Márquez. Post lunch, Gabriel Iturra returned to host a hands-on workshop on "Web Scrapping". The day concluded with collaborative group exercises, allowing participants to integrate their newfound knowledge and share insights with peers.
* **Thursday, July 20, 2023**. The fourth day centered around Visualization and Data Analysis. The day was initiated with a lecture on "Data Visualization" by Cinthia Sánchez. Following a coffee respite, Cinthia continued with an interactive workshop on the same topic. After lunch, attention shifted to a lecture on "Unsupervised Learning" presented by Hernán Sarmiento. A short break later, Hernán led a workshop, delving deep into the intricacies of unsupervised learning techniques.
* **Friday, July 21, 2023**. The event's fifth day adopted a unique format emphasizing "Research Speed Dating" and group formation activities. The day started with a keynote on "Plataforma Telar" by Juan Pablo Luna. Post coffee, Andrés Abeliuk introduced a workshop on the "ChatGPT API". The mid-day session discussed the logistics of the Research Speed Dating. Post lunch, participants engaged in "Research Speed Dating" sessions, first with maximally similar groups and then with maximally dissimilar ones. The activities culminated with a group debrief, where project ideas were deliberated and selected.

## Second Week

During the second week of the event in July 2023, participants balanced intensive group projects with enlightening keynotes and informative sessions. Topics ranged from social innovation in data to the nuances of digital social data interpretation. The week culminated with teams proudly presenting their collaborative endeavors.

* **Monday, July 24, 2023**. The sixth day was characterized by a balance of intensive group project work and informative keynotes. The day commenced with a keynote on "La Ciencia de Datos como Motor de la Innovación Social" (The Science of Data as an Engine for Social Innovation) presented by Jazmine Maldonado from Innovación y Transferencia Tecnológica, IMFD. After a morning coffee break, Marcelo Mendoza captivated the attendees with insights on "Disinformation and network analysis". The afternoon was primarily devoted to group projects. Attendees had substantial blocks of time both before and after a short break to collaborate, brainstorm, and make significant strides on their projects.
* **Tuesday, July 25, 2023**. The seventh day of the event, attendees delved deeply into their group projects, punctuated by an enlightening keynote and a valuable discussion session. The day began with Mauricio Bucca presenting a keynote on "Experimentos online" (Online Experiments). Following a morning coffee interlude, participants plunged into focused project time, allowing teams to progress and refine their ideas. Post-lunch, the afternoon kicked off with another productive project session, this time enriched by the presence of Claudio Gutiérrez, who facilitated an ethics discussion pertaining to the projects. After a brief recess, the day rounded off with further dedicated project time, empowering groups to advance their endeavors.
* **Wednesday, July 26, 2023**. The eighth day of the event, a blend of group work, enlightening lectures, and insightful panels took center stage. The day's proceedings began with a remote keynote by Gonzalo Rivero from Pew Research Center, DataLabs, delivered via Zoom. Following a morning coffee break, an intriguing industry panel was hosted by Pablo Beytía, featuring insights from DESUC with Cristián Ayala and Ulhoster with Antonio Díaz-Araujo. After a lunch break, attendees were treated to a lecture by Pablo Beytía on the "Challenges interpreting digital social data," shedding light on the intricacies of understanding online behaviors and patterns. The remainder of the day was dedicated to group project time, allowing participants to further collaborate and fine-tune their initiatives.
* Thursday, July 27, 2023. The ninth day was marked by concentrated project work and the much-anticipated group presentations. The day kicked off with dedicated project time, allowing teams a final opportunity to refine their work and prepare for the presentations. Following a morning coffee break, participants continued their project endeavors, making last-minute adjustments and rehearsals. After lunch, the event transitioned into a showcase of group achievements, with the first round of group presentations taking the spotlight. A brief intermission allowed attendees to gather their thoughts before they plunged into the second round of presentations, where teams proudly displayed the culmination of their collaborative efforts and insights.

## Other reflections, challenges, and observations.

Reflecting on the SICSS event in Chile, our logistics and facilities primarily catered to the needs of the participants, although we faced certain challenges, especially regarding meals. Our communication efforts, backed by the IMFD communication team, were fruitful, and representation showed diversity in nationality and educational background, even though we observed a concentration of participants from specific demographics.

**Logistics:** The event undoubtedly thrived due to meticulous logistics planning. A significant advantage was securing a well-equipped room from the university, complete with necessary amenities like WiFi and chargers; the winter break ensured the venue's uninterrupted availability throughout the event. However, the first week posed a challenge concerning the provision of lunch. Our initial inability to pay for catering meant we had to organize lunches ourselves. This became quite an ordeal, especially ensuring timely delivery while accommodating diverse dietary restrictions. Things became considerably smoother once we transitioned to catering professional services, highlighting the importance of such services for future events.

**Communication:** For our inaugural event in Chile, the response and interest from students were commendable. The IMFD's communication team was instrumental in enhancing our outreach. Their efforts to produce and disseminate video capsules across platforms like LinkedIn, Instagram, and Twitter expanded our visibility and probably played a pivotal role in attracting a diverse participant pool.

**Representation:** Our endeavor to maintain gender parity was met with varied results. While we ensured a balanced gender representation among participants, achieving the same for speakers and the organizing team proved challenging. An unexpected yet pleasant surprise was the participation of individuals from various educational backgrounds ranging from bachelor students to faculty members and those unaffiliated with Chilean universities or without Chilean nationality. However, a significant portion of the attendees were Chilean, based in the capital, and were amid their Ph.D. pursuits. This indicates the necessity for broader outreach in the future to ensure a more varied representation.

## Further recommendations

From our experiences hosting the event, we have distilled some key lessons and recommendations to aid future organizers. These insights predominantly revolve around two areas: optimizing communication with participants throughout the event and accommodating participants' diverse expertise with the programming language Python.

**1. Communication Channels:**

* **Efficiency and Familiarization:** We utilized a range of platforms, including email, Slack, Google Drive, and our dedicated website. However, we noticed a variance in participants' comfort and proficiency in accessing these platforms. It might benefit future organizers to initiate the event with a brief orientation on navigating these platforms. Furthermore, continually emphasizing the significance and purpose of each platform can encourage more proactive participation. A silver lining, we observed was the collaborative spirit among attendees, with many helping each other access information.
* **Post-event Access to materials:** After the event, we consolidated all the resources into an open-access GitHub repository (<https://github.com/ppseguel/sicss-chile-2023/>). These resources, being in Spanish and Python-centric, supplement areas in which the current SICSS curriculum might be sparse. We kept a Google shared folder for participants with a bit more materials. Yet, for those inspired to adopt a similar approach, obtaining consent from speakers and TAs about sharing materials is essential. Consider that clarity regarding which versions of materials are public and which remain restricted is crucial.

**2. Catering to Varied Python Expertise:**

* **Interpreting Pre-event Skill Assessment:** Our pre-event survey gauged participants' proficiency with Python and Colab. Yet, the depth of these insights truly manifested during the event. Organizers should perceive such surveys as initial indicators and remain flexible, adjusting the program in response to real-time feedback.
* **Feedback Loops and Adjustments:** We adopted an informal approach to gather feedback during the early stages of the event. This helped identify knowledge disparities. For future events, a more structured approach, like simple gestures like raising hands or a quick online poll, can reveal much about participants' comfort levels and understanding.
* **Emphasizing Onboarding:** The survey results shouldn't detract from the significance of onboarding materials. As sessions progress, their complexity might catch some off-guard. Organizers can ensure participants are better prepared by highlighting how onboarding materials dovetail into the live sessions.
* **Flexible Composition of Project Teams:** The composition of project teams should mirror the varied expertise of participants. We focused on topical interest. If it becomes evident that certain members are facing challenges, consider reshuffling groups fast or assist more directly. This promotes a balanced skill set within each team, fostering collaborative learning.

To wrap up, our event showcased the vibrancy and potential of collaborative learning. These lessons, drawn from our successes and challenges, highlight the importance of adaptability and resilience. They can serve as valuable insights to make future events even more effective and participant-friendly.