

Study Protocol and Provenance

Research questions, longitudinal case design, evidence rules, analysis procedures, provenance, privacy, and data availability

Associated manuscript: Training Space Software Engineers: Connecting Computing, Science, and Professional Practice in an International School

Version 1.0.7, review material, 28 August 2026

Replication package DOI: [10.5281/zenodo.22117082](https://doi.org/10.5281/zenodo.22117082)

Prepared by Francesco Basciani, Martina De Sanctis, Franco Raimondi, Daniele Masti, Amleto Di Salle, Martina Feliciani, Davide Zaminga, Riccardo Mastropietro, and Patrizio Pelliccione.

This document contains deidentified study documentation. Participant level administrative records, questionnaire rows, open comments, identity documents, financial information, visa files, private communications, and travel records are not included.

1. Purpose and Scope

This package supports a retrospective longitudinal case study of two successive editions of an international Space Software Engineering school. The unit of analysis is the program transition across the two editions, not individual participants and not a participant panel. The material permits recalculation and audit of released aggregate results.

The study examines one documented transition from 2025 to 2026. It does not represent a controlled effectiveness comparison and does not establish a recurring improvement cycle.

Evidence is kept at four levels: planning rationale, scheduled provision, verified process or delivery records, and participant reactions. Learning, interdisciplinary integration, transfer, durable professional relationships, equitable attendance, and sustainability require direct assessment or follow up and are not inferred from the present evidence.

2. Research Questions

RQ1

How were documented needs, constraints, and opportunities from one edition translated into program changes in the next, while preserving the school's founding educational objectives?

Analytical target: Documented changes and retained founding goals across one observed transition

Direct evidence: Four founding goals, the published first edition evaluation, dated 2025 planning records, alternatives and decisions, agendas, activity designs, exact track choice patterns, and operational role records.

Inference boundary: Only D6, D7, and D11 contain complete process traces. Founding goal compatibility is assessed after reconstructing each chain. The records establish one observed transition and alignment with declared curricular demand, not exclusive causality, optimality, attendance, or a recurring multi edition cycle.

RQ2a

How were disciplinary perspectives, professional expertise, practical activities, responsible innovation, and networking jointly configured across the two editions?

Analytical target: Program configuration across five educational dimensions

Direct evidence: Published program records, agendas, activity designs, exact track choice patterns, track specific expertise prompts and aggregate completeness, source bounded contributor records, practical format coding, communication artifacts, bounded organizer reported communication uses, and a bounded organizer account of Rover laboratory delivery.

Inference boundary: Scheduled configuration, declared track demand, contributor records, expertise prompts, reported communication use categories, and reported laboratory structure do not establish complete delivery, verified group balance, activity breadth, participant exposure, interdisciplinary learning, dialogue quality, mechanism effects, transfer, or persistent relationships.

RQ2b

What participant reactions were reported in relation to this program configuration?

Analytical target: Returned participant reactions to the program configuration

Direct evidence: Complete questionnaire category distributions, verified item definitions, a harmonized item crosswalk, and descriptive reaction indicators for both editions.

Inference boundary: Returned forms describe separate response sets and do not establish unique respondents, population estimates, change, causal effects, learning, or track specific reactions.

RQ3

How was international participation supported from outreach to on site delivery, and which organizational resources and design choices enabled this process?

Analytical target: Service architecture and resources supporting international participation

Direct evidence: Outreach, application, separate track evaluation and admission decisions, preference to admission transitions, scholarship selection and amount formulas, visa assistance, differentiated registration, communication, staffing, logistics, safety, capacity, and weekly delivery records.

Inference boundary: Stages use different units. Separate track decisions and cost adjusted scholarship formulas establish process structure, not scoring reliability, educational quality, equal opportunity, or support adequacy. Longitudinal outcome, efficiency, and sustainability analyses require linked attendance, outcome, effort, and cost measures.

3. Founding Objectives, Emergent Priorities, and Support Condition

ID	Program element	Status	Evidence boundary
O1	Advanced and interdisciplinary Space Software Engineering training	founding goal	Curricular sequence, representation, and declared track demand do not establish instructional quality, attendance, interdisciplinary competence, or learning.
O2	Integration among academia and professional organizations	founding goal	Contributor records and agenda opportunity do not establish joint curriculum design, verified delivery, dialogue quality, collaboration, or educational effect.
O3	International learning environment	founding goal	Profile and process reach do not establish inclusion, completion, unique attendance, cross cultural learning, equitable participation, cost sufficiency, or affordability.
O4	Responsible innovation and knowledge or technology transfer	founding goal	Content and questionnaire reaction do not establish competence, changed practice, completed transfer, or long term viability.
EP1	Mission oriented practical experience	emergent priority	EP1 is an emergent design priority, not a founding objective or independent validation criterion; planning design and scheduled provision do not establish complete delivery, verified group balance, participant exposure, artifact quality, learning, or successful reuse across audiences.
EP2	Structured face to face dialogue and networking	emergent priority	EP2 is an emergent design priority, not a founding objective or independent validation criterion; provision, reported use categories, and immediate reactions do not establish activity breadth, mechanism specific effects, inclusion, tie quality, or persistent relationships.
S1	Support for international participation in person	support condition	S1 is not an educational outcome or evidence of complete process outcomes, adequacy, efficiency, total cost, or sustainability.

4. Analysis Procedures

RQ1, traceable program change

- Recover O1 to O4, initial decisions, published lessons, and identified needs from the first edition.
- Order later planning, agenda, communication, and administrative records by date.
- Include a change only when a dated need, constraint, or opportunity, a later alternative or decision, and a corresponding later program or organizational record are all present.
- Reconstruct the adaptation chain before assessing compatibility with O1 to O4.

- Use EP1 and EP2 to describe priorities emerging from first edition evaluation, not as retrospective validation criteria.
- Preserve alternative explanations such as recruitment, funding, contributor availability, and institutional opportunity.
- Reconstruct exact 2025 entry paths from published aggregate overlaps and derive exact 2026 paths from unique applicant identifiers.
- Treat track choices as declared demand after portfolio announcement, not as validation of the portfolio decision, attendance, or interdisciplinary learning.

RQ2a, program configuration

- Extract disciplinary perspectives, professional expertise, practical activities, responsible innovation, and networking provision from retained program records.
- Preserve source taxonomies and report agendas, governance, sequence, and track choices as program design and declared demand, not as proof of learner integration.
- Classify a timetable slot as practical only when its title or description explicitly identifies a tutorial, demonstration, laboratory, practical activity, or facilitated working group.
- Extract the separate Robotics, AI, and GEO expertise prompts and aggregate response completeness without releasing applicant text or treating self reports as validated competence.
- Treat the six macro groups, collaborative lunar scenario, nominal points, and use of Robotics expertise in group composition as a bounded organizer account supported by the surrounding agenda, activity design, contributor, and application records.
- Treat participant travel, accommodation, and room sharing exchanges and organizer notices, materials, and slide distribution as bounded organizer reported use categories. Without message or activity logs, do not infer use volume, participation breadth, tie formation, or persistence.

RQ2b, participant reactions

- Report complete questionnaire categories by edition and summarize common items using the declared item level crosswalk.
- Do not calculate effects, significance tests, or questionnaire results by track.

RQ3, international participation support

- Reconstruct the recorded path from outreach and application through selection, scholarships, visa support, registration, communication, and delivery.
- Treat each applicant by selected track as the admission decision unit and derive preference to admission multiplicities from the separate track statuses.
- Report every quantity with its original source unit and administrative stage.
- Keep scholarship requests, selections, allocations, and payments separate.
- Reconstruct the scholarship amount rule from the country allowance table, the weekly accommodation constant, and the formulas used across track selections.
- Treat communication, staffing, logistics, safety, and capacity as service provision, not as educational outcomes.
- Do not infer scoring reliability or educational quality from separate track decisions, and do not estimate equity, efficiency, affordability, cost effectiveness, or a complete participation funnel without linked outcomes and comparable resource data.

Track sequence evidence

The table records exact and mutually exclusive entry paths for the published 2025 registration aggregates and the 2026 applicant preferences. In 2025, 62 of 66 people selecting several tracks included AI. In 2026, 194 of 197 applicants selecting several tracks included AI. These patterns characterize declared demand for curricular paths.

Edition	Stage	Path	Position	Count	Denominator
2025	entry registrations	ROB	single track	22	153
2025	entry registrations	DT	single track	16	153
2025	entry registrations	AI	single track	49	153
2025	entry registrations	ROB and DT	adjacent tracks	4	153
2025	entry registrations	DT and AI	adjacent tracks	11	153
2025	entry registrations	ROB and AI	non adjacent tracks	6	153
2025	entry registrations	ROB and DT and AI	complete sequence	45	153
2026	applicant preferences	ROB	single track	42	444
2026	applicant preferences	AI	single track	68	444
2026	applicant preferences	GEO	single track	137	444
2026	applicant preferences	ROB and AI	adjacent tracks	63	444
2026	applicant preferences	AI and GEO	adjacent tracks	35	444
2026	applicant preferences	ROB and GEO	non adjacent tracks	3	444
2026	applicant preferences	ROB and AI and GEO	complete sequence	96	444

Track level selection evidence

Applicants could select several tracks, but every selected track received a separate evaluation and admission decision. The resulting preference to admission transitions show that admission to one track did not transfer automatically to another. Of the 96 applicants selecting all three tracks, 26 were admitted to one, 37 to two, one to all three, and 32 to none.

Selected path	Admitted tracks	Applicants	Path denominator
ROB and AI and GEO	0	32	96
ROB and AI and GEO	1	26	96
ROB and AI and GEO	2	37	96
ROB and AI and GEO	3	1	96

Cost adjusted scholarship amount

The 2026 allocation rule combined a country indexed travel component with a common weekly accommodation component. The travel component was counted once for recipients selected for several tracks, while the accommodation component was repeated for every awarded week.

Component	Basis	Frequency	Amount or range
Country indexed travel allowance	Average return airfare between Italy and the capital of the applicant's country of residence	Once per scholarship recipient, including recipients awarded more than one track	EUR 100 to EUR 1300, 70 countries
Weekly accommodation allowance	Common average room cost in L'Aquila for one week	Once for each awarded track week	EUR 350
Multiple track composition	Travel allowance included once and weekly accommodation added for each awarded track	Applied across all awarded tracks for the same recipient	Composition rule

The records establish the allocation rule and its geographic adjustment. They do not establish actual travel expenditure, cost sufficiency, award acceptance, payment completion, attendance, or equal participation opportunities.

5. Rover Systems Lab Evidence

The laboratory evidence combines the official full day agenda, a retained design record, a contributor roster, source derived application fields, and a bounded organizer account of delivery. These layers establish different parts of the activity and are not treated as interchangeable.

Element	2026 record	Evidence status	Boundary
Mission framing	A simplified lunar resource logistics scenario connected simulated rover behavior, obstacles, autonomy, and mission constraints	mixed artifact report	The design record and organizer account establish the scenario but not participant exposure task completion or learning
Simulation and common platform	GSSI Robotics Laboratory instructors and international contributors explained and simulated the rover system using a common technical environment	mixed artifact report	The retained sources establish design and reported delivery but not setup quality actual timing or group outcomes
Activity sequence	Mission briefing, instructor simulation, system setup, integration, control, testing, demonstration, and design review	mixed artifact report	The design sequence and reported delivery do not establish complete execution artifact quality or learning
Intergroup cooperation	Six macro groups coordinated simulated rovers toward one collective objective of moving resources in the lunar environment; obstacles and nominal points represented task constraints	mixed artifact report	The six group structure is organizer reported; nominal points were not an assessment measure and no task completion or coordination observations are retained
Expertise informed grouping	The application collected separate Robotics, AI, and GEO skill entries; organizers report using Robotics entries to distribute prior experience across the six macro groups	source derived field and organizer report	The prompts and aggregate completeness are source derived but no group assignment matrix or competence validation was retained
Organization	Multiple planning meetings, four ninety minute agenda slots, and seven contributors from academic and industrial organizations supported the full day laboratory	mixed artifact report	Meeting count and person hours are not released and scheduled contributors do not establish complete delivery
Training identity	Organizers treated the 2026 delivery as the initial SPACERAISE Academy laboratory format	organizer report and documented reuse intention	The delivery characterization is organizer reported and the reuse path is an intention not evidence of successful transfer across audiences

Track specific expertise fields

The application requested track specific self reports. The Robotics entries were used by organizers as an input when distributing prior experience across six macro groups. No participant text is released.

Track	Prompt	Selected	Nonplaceholder	Missing	Reported use
ROB	Please list the robotic software you are familiar with (e.g., ROS, Python, etc.)	204	183	21	Organizers report using the Robotics entries to distribute prior experience across six macro groups.
AI	Please list the AI software you are familiar with (e.g., Python, TensorFlow, etc.)	262	227	35	Available as track specific self reported input; no group composition use is claimed.
GEO	Please select the software you are familiar with	271	270	1	Available as track specific self reported input; no group composition use is claimed.

The released records do not reconstruct group membership, validate competence, measure balance, record actual timing or task completion, preserve produced artifacts, or evaluate cooperation and learning.

6. Networking and Communication Evidence

The communication evidence separates infrastructure, reported use, immediate participant reactions, and outcomes requiring follow up. Discord opened about one month before delivery and contained general, scholarship, material, and track specific spaces. Organizers report participant use for travel, accommodation, and room sharing and organizer use for notices, speaker materials, and slides.

ID	Need	Implemented feature	Observed or reported evidence	Boundary
C01	Organizer-stated need for pre-arrival travel, accommodation, scholarship, and general-logistics coordination without requiring phone-number exchange.	A channel-based communication server opened about one month before the event; email remained the primary official channel.	Retained channel taxonomy plus organizer-reported participant use before arrival to exchange travel and accommodation information and arrange shared hotel or bed and breakfast rooms.	The sources establish available spaces and reported use categories, not message volume, active-user reach, reciprocity, privacy outcomes, community formation, or continued contact.
C02	Organizer-stated need to distribute speaker materials and track-specific information through persistent topic spaces.	Materials and track-specific channels supplemented official email.	Retained channel taxonomy plus organizer-reported use for logistical notices, speaker materials, and slides.	No message export, channel activity log, delivery receipt, or readership measure was retained.
C03	Create low-stakes opportunities for participant interaction outside formal talks.	Planned networking/social sessions, including one app-supported group sensory activity designed around a collaborative task.	Official agenda and activity artifact.	The artifact documents the designed format, not contact formation, collaboration quality, causal effect, or durable relationships.

The available evidence supports reported categories of use and immediate questionnaire reactions. It does not measure message volume, active participation, network structure, relationship quality, or persistence. CacaoConnect is retained as a designed social format, not as evidence of networking effectiveness.

7. Documented Transition Traces

A trace is retained only when the material contains a prior dated basis, a subsequent alternative or decision, and a corresponding later agenda or organizational record. Three traces satisfy this rule. Founding objective compatibility is assessed separately after the chain has been reconstructed. A trace documents a decision pathway; it does not establish exclusive causality, optimality, or educational improvement.

Driver	Documented basis	Alternatives and decision	Implemented change
D01	The 2025 participant and speaker evaluation requested greater practical depth, before planning for the second edition.	Continue distributing short practical sessions, or concentrate time and resources in recognizable mission activities; September planning required more practice and advance confirmation, and the Rover design expanded from a half day concept to a full day.	The 2026 agenda schedules eight named formats across 11 timetable slots, including a full Rover day, an AI working group, and GEO laboratories; organizers report six interdependent Rover groups working on one lunar logistics objective.
D02	The September 11 planning record identifies successful Earth Observation content, difficulty assembling a complete Digital Twins speaker portfolio, and the separation of Robotics and AI by Digital Twins as a barrier for interested participants.	Planning moved from a possible two week Robotics and AI format to a three week progression; by October 10, Digital Twins content was retained within Robotics and GEO was defined around spatial data, Social Science, practical work, and an explicit bridge from AI.	The final agenda progresses from physical and software systems in Robotics, through computational methods in AI, to societal analysis of space data in GEO; Social Science owns GEO while Digital Twins remains represented within Robotics.
D03	The first edition reported coordination and logistics as organizational challenges before second edition staffing decisions.	Continue shared informal responsibility, or dedicate staff to participant and complex administrative work; October planning proposed one participant resource and then one or two additional people.	The 2026 role register separates participant management, scholarships, reporting, teaching administration, project management, and delivery.

8. Prospective Evaluation Protocol

The following protocol identifies evidence required to test claims that the current retrospective case cannot establish. It is a future evaluation design, not a set of validated outcomes.

ID	Evaluation question	Unit and timing	Evidence to collect	Success indicator
P2	How do participants integrate disciplinary perspectives?	Participant or group, before and after an integrated task	Shared space problem, submitted artifact, explicit integration rubric, independent scoring, and prior background.	Demonstrated integration of concepts and evidence across domains.
P4	How do practical formats support learning and intergroup coordination?	Participant, group, and interdependent group system during each laboratory	Participant roster, group assignment matrix, prior expertise fields, actual timing, exposure duration, facilitator and equipment load, task completion log, cross group coordination observations, submitted artifacts, task specific rubric, and activity feedback.	Artifact quality gains on task specific outcomes and observed coordination across interdependent groups.
P3, P4	How does thematic grouping support expert dialogue?	Session, observed during delivery	Delivered roster, scheduled and actual discussion time, structured observation of questions and exchanges, and participant item.	Sustained questions, exchanges, and connections across contributions.
P4	How does networking provision support professional relationships?	Consenting participant and privacy preserving platform aggregate, before, during, after, and at timed follow up	Channel activity aggregates by period and channel, unique active accounts, activity participation, inclusion indicators, consented contact network, new tie formation, and later contact or collaboration.	Broad participation in the intended communication uses and new ties followed by later contact, exchange, or collaboration.
P5	How does participation support enable attendance?	Applicant, through every administrative stage	Stable pseudonymous identifier, selection criteria assessment, decision, visa outcome, payment outcome, registration, attendance, delay, and exit reason.	Completion of support stages and attendance across participant profiles.
P5	How sustainable is the delivery model?	Edition and service process	Staff hours by role, service volumes, resolution times, final budget by comparable category, incidents, and unique attendance.	Reliable service delivery within planned capacity, effort, and budget.
P1, E1	How does a program change affect its target outcome?	Decision and affected participants, specified before delivery	Dated rationale, alternatives, intended mechanism, implementation check, baseline or comparator, and preregistered outcome.	Improvement on the predefined outcome with confirmed implementation.

9. Data Provenance

The public workbook contains aggregate representations only. The source aliases below identify the origin and custody of each released quantity without disclosing participant identities or private files.

ID	Edition	Source alias and type	Custody	Public representation
S01	first	published_first_edition_study, published article and associated aggregate artifacts	public bibliographic source	aggregate rows in administrative-layers track-selection participant-profiles and questionnaire-binary-indicators
S02	second	application_intake_workbook, restricted administrative workbook	institutional custodian	application submissions and normalized email-key counts
S03	second	decision_and_selection_workbook, restricted administrative workbook	institutional custodian	applicant IDs track preferences separate Robotics AI and GEO expertise fields aggregate expertise response completeness admissions scholarship requests and selections visa requests exact process intersections weighted scoring fields decision-stage country counts country indexed travel allowances weekly accommodation constant and cross track scholarship formulas

ID	Edition	Source alias and type	Custody	Public representation
S04	second	registration_workbooks, restricted administrative workbooks	institutional custodian	registration submissions valid email-key counts required-field presence and exact email linkage aggregates
S05	second	badge_workbook, restricted operational workbook	institutional custodian	badge-row and normalized-name counts
S06	second	capacity_planning_artifacts, restricted operational notes and room-capacity artifact	institutional custodian	aggregate operating cap and approximate largest-room capacity
S07	second	participant_questionnaire_workbook, restricted questionnaire response workbook	institutional custodian	closed-ended aggregate distributions corpus-size metadata and reconstructed headers
S08	second	invited_speaker_workbook, restricted operational workbook	institutional custodian	accepted-speaker counts split by track and source-provided binary industrial field
S09	second	official_agenda_snapshot, public web snapshot retained by authors	author workspace	author-curated deidentified agenda summary and practical-slot classifications
S10	first	zenodo_first_edition_participant_workbook, public questionnaire response workbook	Zenodo record 18416338	aggregate discrepancy check only; workbook is not mirrored
S11	first	scholarship_allocation_minutes, restricted signed allocation and reallocation minutes	institutional custodian	final aggregate scholarship allocation amount and derivation
S12	second	september_2025_planning_record, restricted contemporaneous meeting summary and transcript	author workspace	sanitized redesign drivers in rq1-theme-frame and redesign-traceability
S13	second	october_2025_planning_record, restricted contemporaneous meeting summary and transcript	author workspace	sanitized redesign drivers in rq1-theme-frame and redesign-traceability
S14	second	scholarship_criteria_document, restricted administrative PDF	institutional custodian	criterion names compared with operational workbook fields
S15	second	administrative_staffing_planning_records, restricted contemporaneous meeting summaries and transcripts	author workspace	sanitized redesign driver and staffing decision
S16	second	rover_systems_lab_design_record, restricted activity design record	author workspace	sanitized mission framing common platform simulation basis activity sequence and Academy reuse path in rover-systems-lab-design
O01	second	broad_demand_organizer_account, organizer reconstruction	no record-level dataset	qualified aggregate only
O02	second	capacity_and_headcount_organizer_account, organizer reconstruction	no final attendance register	qualified cap and weekly headcounts
O03	second	communication_infrastructure_and_use_account, organizer reconstruction plus retained screenshot	author workspace	approximate membership, deidentified channel taxonomy, participant prearrival coordination uses, and organizer information and material distribution uses
O04	second	operations_staffing_organizer_account, organizer reconstruction	no time-allocation log	qualified lower-bound count only
O05	second	interaction_design_organizer_account, organizer reconstruction plus agenda structure	author workspace	qualified thematic grouping and protected discussion rationale
O06	second	safety_and_opening_organizer_account, organizer reconstruction plus restricted operational artifacts	institutional custodian	qualified aggregate safety provision incident and opening scale context

ID	Edition	Source alias and type	Custody	Public representation
O07	second	rover_systems_lab_organization_account, organizer reconstruction plus retained design agenda contributor and application field records	author workspace	qualified organizational process reported six group collaborative lunar scenario and expertise informed group composition combined with scheduled duration contributor count and source derived expertise field aggregates

10. Privacy and Deidentification

- All public participant data are aggregate.
- Names, email addresses, identifiers, source linkage keys, timestamps, and free text are excluded.
- Country information is released as counts, not as participant level country lists.
- Speaker names and institution names are excluded from the agenda classification.
- Organizer estimates retain qualifiers such as about, almost, and more than.

The school name is omitted, but the case may remain recognizable from public contextual information and authorship. The package is therefore deidentified, not anonymous.

11. Ethics and Data Availability

An institutional ethics determination has been obtained for this study. The exact responsible body, determination type, date, protocol identifier if applicable, and publication reuse scope must be transcribed from the retained institutional record before public archival release.

The deidentified replication package is archived on Zenodo under DOI 10.5281/zenodo.22117082.

Participant level administrative and questionnaire records are not released because they contain personal or potentially sensitive information. Identity documents, banking and tax information, visa material, travel records, private communications, and verbatim open responses remain under institutional custody.

12. Released Files

The web package contains three human readable documents together with machine readable data and executable analysis artifacts:

- This study protocol and provenance document.
- The questionnaire instruments and response definitions document.
- The aggregate data and analysis workbook, including an index, data dictionary, program goals, decisions, traceability, questionnaire distributions, source schemas, and released results.
- Canonical CSV tables, Python analysis and validation code, exact LaTeX table and figure sources, generated SVG figures, and JSON validation results.

The analysis build currently passes 253 of 253 released arithmetic, schema, provenance label, privacy, and internal consistency checks. A pass does not independently validate author coding against restricted source artifacts.