DESTINI

SMART DATA PROCESSING AND SYSTEMS OF DEEP INSIGHT

http://www.destini2020.eu

DESTINI

Smart Data Processing and Systems of Deep Insight

Deliverable 5.2

Report on best practices for Supervision, Mentoring and Tutoring

Document details:

Editor:	JADS
Contributors:	CUT, UNIROMA
Date:	10/10/2022
Version:	8.0

Document history:

Vers.	Date	Contributor	Comments
1.0	18/06/2022	D.A. Tamburri	Initial document, structure and content
2.0	21/07/2022	W.J.A.M. vd Heuvel	Additions and corrections
3.0	05/08/2022	A.S. Andreou Partners	First review
4.0	07/09/2022	S. Mappouras	Additions and corrections
5.0	05/09/2022	A.S. Andreou, W.J.A.M. vd Heuvel	Final review
6.0	30/09/2022	Partners	Final review & approval
7.0	10/10/2022	A.S. Andreou	Final additions and corrections
8.0	10/10/2022	Partners	Final review & approval

Contents

1. Introduction

1.1 Purpose

The purpose of this deliverable is to offer details on the practices followed and plan templates prepared to expedite the Ph.D. trajectories enacted in the scope of the DESTINI action, as supported by partners at JADS, CUT, and UNIROMA (Sapienza). The three partners bootstrapped this activity with a brainstorming and knowledge exchange activity and workshop at the very beginning of Y1 to determine (1) what previous best practices were institutionalized at an organizational level in their respective universities and (2) what best integration approach was possible. The deliverable provides an outline of the practices that were used for coordinating research of PhD candidates at the three participating universities and aligning it with the areas of interest to DESTINI and for mentoring and tutoring Early-Stage Researchers that participated in the project's mobility program.

1.2 Definitions, Acronyms, and Abbreviations

CUT: Cyprus University of Technology

RPA: Robotic Process Automation

CRISP-DM: Cross-industry Standard Practice for Data Mining

ML: Machine-Learning

AI: Artificial Intelligence

DevOps: Software Development and Operations

ESR: Early-Stage Researchers

JRA: Joint Research Areas

1.3 Overview

The rest of the document is structured as follows: Section 2 and its associated sub-sections include material featured to address studentship training setup and research design and execution, as well as basic principles that are applied to the supervision of PhD candidates that will be respected and followed for ESRs. Section 4 sketches the approach employed for mentoring and tutoring ESRs that participated in DESTINI's mobility program. Finally, Section 4 concludes the document.

2. Research Design and Training Setup: General Approach and Key Challenges

2.1 Foundations for Ph.D. Supervision reflected on ESRs

The standard approach for training supervision setup intended for the collaborations stemming and active within the boundaries of the DESTINI project draws from mainly two sources, (1) the training and supervision guide enacted by the Stichting Katholieke Universiteit Brabants—Tilburg University—unit of JADS, part of JADE lab, as well as (2) the Cross-Industry Standard Process for Data Mining (CRISP-DM) guide.

On the one hand, the Tilburg university guidelines entail a reflection over the various Tilburg schools of thought (e.g., see here: https://www.tilburguniversity.edu/education/phd-programs/regulations).

On the other hand, such regulations and reflections were augmented with the CRISP-DM approach, a unified industrial standard introduced in the late 90's and featuring — quoting from its standard guide —"*a solid base for any data-scientific process. [The process] has six sequential phases: (1) Business understanding – What does the business need? (2) Data understanding – What data do we have / need? Is it clean? (3) Data preparation – How do we organize the data for modeling? (4) Modeling – What modeling techniques should we apply? (5) Evaluation – Which model best meets the business objectives? (6) Deployment – How do stakeholders access the results?*". The process in question is recapped in the process model provided in Figure 1, which represents the phases as they relate to each other from a data and scientific flow perspective.

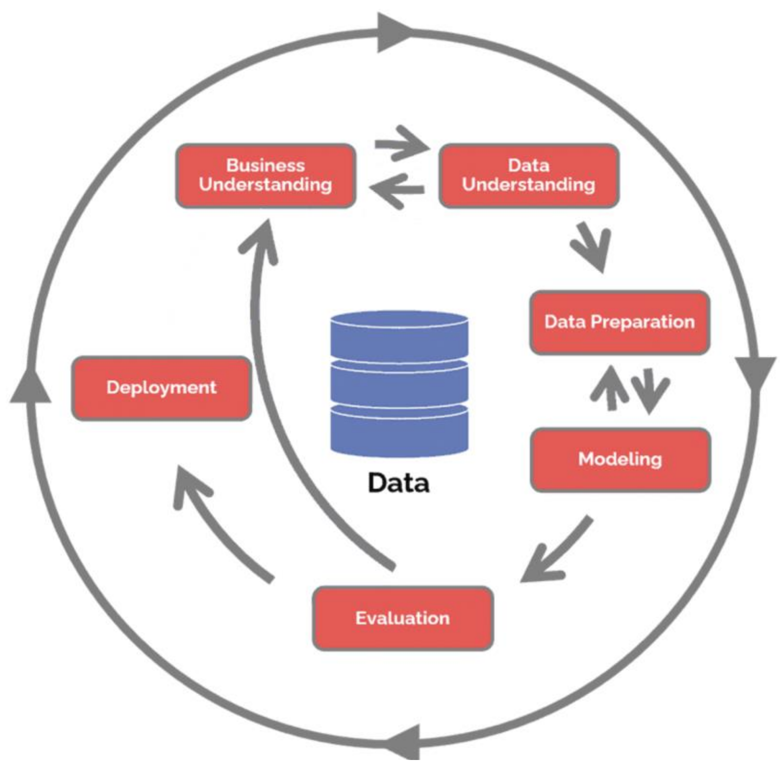

Fig. 1. The Cross-Industry Standard Process for Data Mining (CRISP-DM) Model

In line with this approach, DESTINI scientific discourses followed precisely the six steps reported in the CRISP-DM process and expressed as follows:

1. **Research context and problem understanding**. This phase typically holds the Ph.D. candidate responsible for a fundamental investigation of the problem context in terms of literature available to address it and/or previously used to tackle problems similar in the same domain. This phase is then internally followed by a formalization step, in which the problem context and its characteristics are rendered in a mathematically reasonable form, which limits any bias for further treatise and increases publishability, as well as general internal and external validity. This phase typically concludes with the determination of research questions to be addressed in the scope of the target research activity.

2. **Research Data Requirements and Characteristics Assessment**. This phase aims at determining what data is available for the variables and target improvement factors elicited and formalized in step 1, with the intent of determining also data goals to be further addressed; for example, this phase may entail determining what data mining needs to be performed to investigate a specific Software and/or Data Engineering challenge.

3. **Research and Problem Data Preparation**. Stemming from step 2, this step aims at rendering available data in a format which is usable within and for the scope of construct design and experimental operation; for example, the data mining activities in step 2 are then followed by data wrangling, cleaning, integration, augmentation procedures aimed at making data available for further analysis and synthesis to be conducted in line with research questions definition and their practical experimentation.

4. **Research Analysis and Synthesis**. This phase aims at performing — through a sound experimental design and execution, with sanity checks performed following validity threat assessment approaches — sound research data analysis and results synthesis (e.g., via plots, illustration, descriptive stats and more), to systematically address all research questions defined in steps 1 and 2. This step typically concludes with a brainstorming session over the discussion of obtained results as well as a discussion of further research problems emerged during results elicitation and discussion.

5. **Research results evaluation, confirmation, counterproof, and consolidation/discussion**. This step aims at confirming or double proving obtained results and their discussion, typically involving practitioners and/or other types of so-called domain experts which are able to disprove or confirm/reinforce the findings connected to the discussion of research results in their proper context of operation and impact. This phase is typically supported by qualitative or mixed-methods research approaches.

6. **Results Practice Discussion and Stakeholder Engagement**. Finally, to reinforce the external validity even further and understand the extent of such validity, the results are then ported in a similar context as that introduced and well-formed in steps 1 and 2, and a controlled experiment is enacted through a replication study. Results are typically confirmed or disproved to identify further work on the subject matter.

Each of the six steps also entails parallel revision and peer-confrontation moments of the DESTINI project which were, in one way or another, reinforced through partner collaboration and/or intra-project peer-review; for example, all workshops within the scope of the DESTINI project featured peer comparison and presentation discussion sessions in all research designs, regardless of their state in maturity according to the technology readiness model defined by the European Union. Within such discussion project results and potential exploitation, avenues were identified, discussed, and matured further.

2.2 Training Setup

The main message in this briefing is that empirical software engineering research part of the DESTINI initiative should focus around "*using qualitative empirical research to better support findings obtained in quantitative studies, by providing concrete cases, rationale, and, ultimately, approach causality*".

2.3 Research Design

A typical process of mixing quantitative and qualitative approaches therefore entails, at least:

1) **devising a quantitative study** where qualitative data sources are available as well – for example, analyses of open-source repositories can focus on online communities for which close contacts to community members can be maintained or wherefore the community itself manifests as open and welcoming external inquiry.

2) **planning that quantitative study** to devise, upon analysis, the qualitative investigation of relevant dimensions – for example, phrasing a research question to include the human confirmatory component can aid in identifying the factors that can be controlled and the design of the subsequent qualitative research design.

3) **identifying qualitative means** that can compound the quantitative study and, vice versa, identifying quantitative means that can confirm the qualitative study - for example, while mining software repositories towards hypothesis testing, triangulation of developer interviews for mined

repositories can be planned; at the same time, quantitative statistical measures (e.g., Krippendorff's Alpha coefficient) of coding agreement can be put in place to confirm the validity of qualitative analyses;

4) **discussing identified threats to validity** with a focus on the areas of the study design in which the use of qualitative confirmatory means was not possible.

5) **elaborating follow-up research questions** stemming from the qualitative analysis, such that the theory can be enriched incrementally through further study.

Further on, the section will detail how mixed-methods research plays a role in the aforementioned steps of research operation as well as discussing ways in which such data can be combined effectively for an appropriate scientific discourse to emerge.

2.3.1 Steering Mixed Methods Research within DESTINI

We observed that steering mixed-methods research can be a pain since many factors and variables may emerge constantly during study evolution. We found that using mind-maps where nodes identify variables under study while edges identify type and instance of research methods, helps considerably in:

(a) keeping aligned the study with its quantitative/qualitative connotation and evaluating a balance between the two - for example, a simple edge-counting exercise allows this balance to be checked constantly;

(b) identifying "gaps" in the study for discussion and further analysis - for example, there may be areas of the mind-map that are under-explored and may require further elaboration;

(c) elaborate threats to validity in the context of mixed-methods research - theoretically the entire mind-map should be "covered" in a proper threat to validity discussion.

The figure below recaps the aforementioned interplay as an example of the methodology enforced in the scope of the DESTINI action:

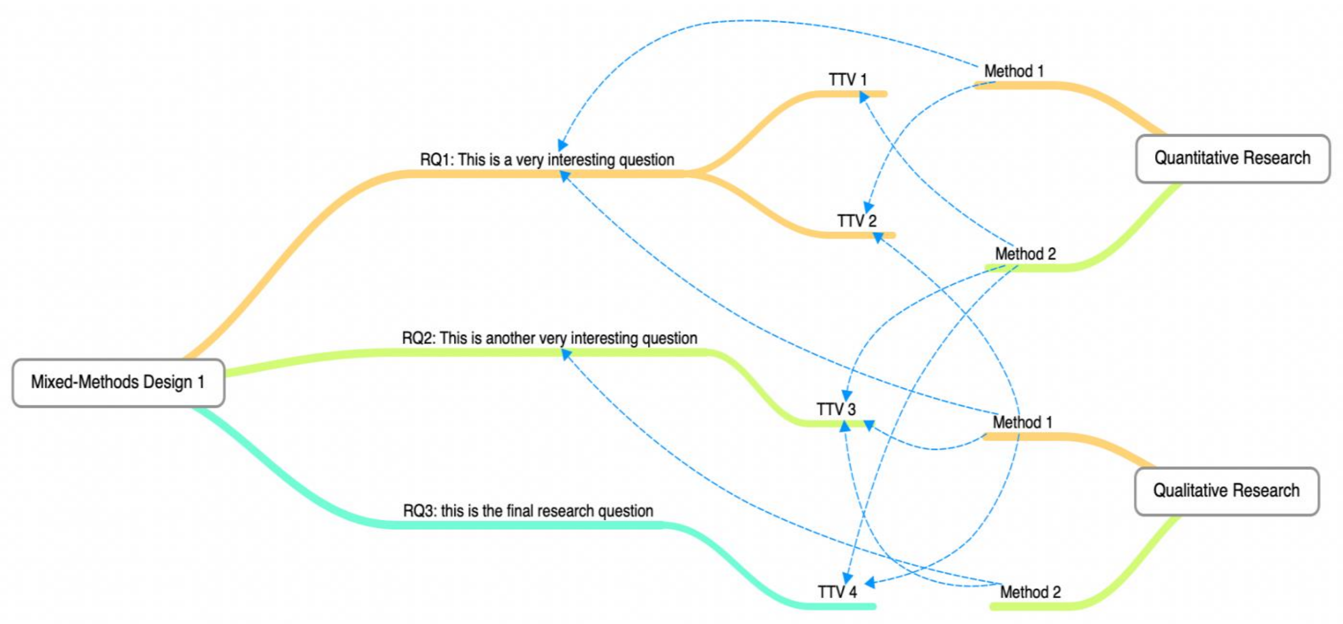

Fig.2: Mixed-methods design through research questions

2.3.2 Major Qualitative Research Discourses for External Validity Reinforcement

After having outlined the limitations of purely quantitative analysis, and highlighted the potential of mixed-methods, the technical briefing provides an overview of selected qualitative approaches:

1) Grounded-Theory: look at data source (e.g., interview logs, field notes); look for indicators of categories in events and behavior - name them and code them on document; compare codes to find consistencies and differences; consistencies between codes (similar meanings or pointing to a basic idea) reveals categories; memo on the comparisons and emerging categories; eventually category saturates when no new codes related to it are formed; eventually certain categories become more central focus (i.e., the core of the theory);

2) Matrix Analysis: essentially, structure observed qualitative phenomenon using flow-chart or other behavioral diagrams; offers an outline of generalized causation - sometimes implementation is known as Formal Concept Analysis;

3) Event Analysis/Ethnography: emphasis is on finding precise beginnings and endings of events and their characteristics; first step is finding specific boundaries and things that mark boundaries or events; in software engineering, especially useful for online data-sources and/or studies that relate to open-source contribution; after event or phenomenon boundaries are found, find phases in events by repeated viewing ;

4) Hermeneutical Analysis: *hermeneutics = making sense of a written text*; not looking for objective meaning of text, but meaning of text for people in situation; try to bracket yourself out in analysis; tell their story (not yours) using their words; use of context (e.g., time and place of writing) to their view of text; possible context uses in software engineering are organizational culture, situation, and history.

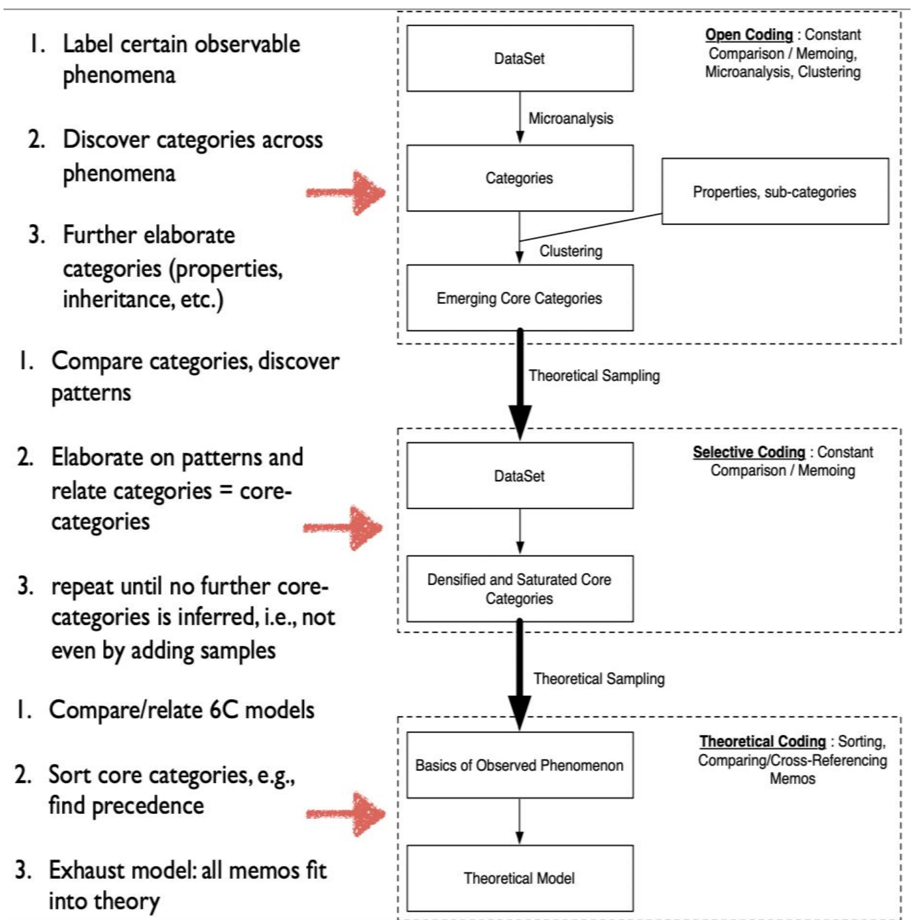

Fig.3: An example of qualitative research

3. Mentoring and Tutoring Approach for DESTINI ESRs

3.1 Mobility program

DESTINI has launched a mobility program targeting at engaging ESRs, mainly from the participating countries, and promote research collaboration in the areas of interest. Mobility of ESRs from the leading universities Sapienza and JADS to Cyprus involved short-scale visits of the order of weeks. These visits were performed at CUT's premises and aimed at investigating specific research challenges that fall within the scope of DESTINI'S JRAS.

The approach followed for mentoring and tutoring ESRs participating in mobility activities are described in this section, building on top of widely known and well-established practices for supervising PhD students, such as those described in section 2.

3.2 Practices for Mentoring and Tutoring

The approach followed in DESTINI for devising a robust, self-contained and efficient set of practises for effectively mentoring and tutoring ESRs was structured in distinct phases as follows:

Phase 1: Definition of stakeholders

The first phase involved the definition of the stakeholders in the corresponding activities. Therefore, two categories were identified: (a) Those resources that acted as mentors/tutors, and, (b) The recipients or beneficiaries of mentoring and tutoring activities. In general, all faculty members of the consortium undertook and played the role of mentor/tutor, supported also by post-doctoral students participating in the project. All resources engaged were experienced researchers with long history in supervising PhD students and conducting research in general. The second category included ESRs from Sapienza and JADS visiting Cyprus, as well as ESRs from CUT that collaborated with the former. All ESRs at the time of the launch of the mobility program were active PhD students, had research agendas aligned with DESTINIs areas of interest, and were mature enough to investigate topics other than those in their PhD proposals but very close to their line of research. Special attention was given to offer equal opportunities to both genders and also to different origins or ethnicity. Therefore, female ESRs were encouraged to participate, and we succeeded in having 3 out of the 8 ESRs relocated in Cyprus to be female researchers.

Phase 2: Relocation Support

This phase was involved with important logistics aspects. Housing was the first and foremost issue that was resolved, with apartments being identified upfront and in close proximity (walk or bus connection) to CUT's premises. Also, all paperwork for the financial support of the ESRs for relocation in Cyprus was prepared and local administration, financial and legal processes were explained to the visiting ESRs in detail.

Phase 3: Set-up proper environment

Special attention was given to setting up a proper working environment of the collaborating ESRs. In this respect, office space was secured upfront with access to infrastructure (e.g., network and Internet, printing facilities and CUT's intranet), and handling entrance passes. The office environment offered very good working conditions, with air-conditioning, silence, access to whiteboards and projectors, telephone lines etc.

Phase 4: Preparatory work

This phase consists of all steps towards best practices for facilitating collaboration between ESR teams and conducting research:

Step 1 – Establish good and effective communication channels. This includes performing regular meetings (physical presence and teleconferencing), planning each meeting (whom, when, what) for sharing knowledge. This step is extremely important as it lays the foundations to create the necessary background for aligning research agendas between ESRs and DESTINI.

Step 2 – Set the time frame and be realistic about what is feasible to achieve within this period and how it may become sustainable and offer a long-term collaboration

Step 3 – Explain the goals of the program, describe research excellence, share the expectations of the consortium and of the ESRs

Step 4 – Transfer and share knowledge through various activities such as:

a. Short scale teaching
b. Presentation of previous published or on-going research work
c. Secure visibility of work and discussion
d. Demonstrate tools and methods developed
e. Reserve time slots for Questions and Answers

Phase 5: Establish and follow a research roadmap

This is the core research phase where a research roadmap is established and followed. To this end, a number of distinct stages are executed, with possible iterations between stages where necessary (see Fig.4) as follows:

Stage - A : Brainstorm – Frequent brainstorming sessions will lead to the selection of the topic(s) to focus on. ESRs are called to discuss their work thus far presented in Phase 4 and identify common areas of interest, such as challenges in their line of research that may be tackled by outcomes of the project (tools, methods, techniques, algorithms) and vice-versa.

Stage - B : Set Targets – For each topic of interest, set research targets to tackle specific challenges.

Stage - C : Outline Novel Idea(s) – Describe briefly idea(s) for solution to the problem(s) to be dealt with.

Stage - D : Perform Background Work – Survey the literature and perform groundwork for an initial investigation of the area in which each challenge and idea falls to provide feedback

Stage - E : Define Research Proposal – Describe a clear methodology and roadmap to achieve research targets.

Stage - F : Produce Documentation – Provide a document in the form of ongoing research or work-in-progress paper.

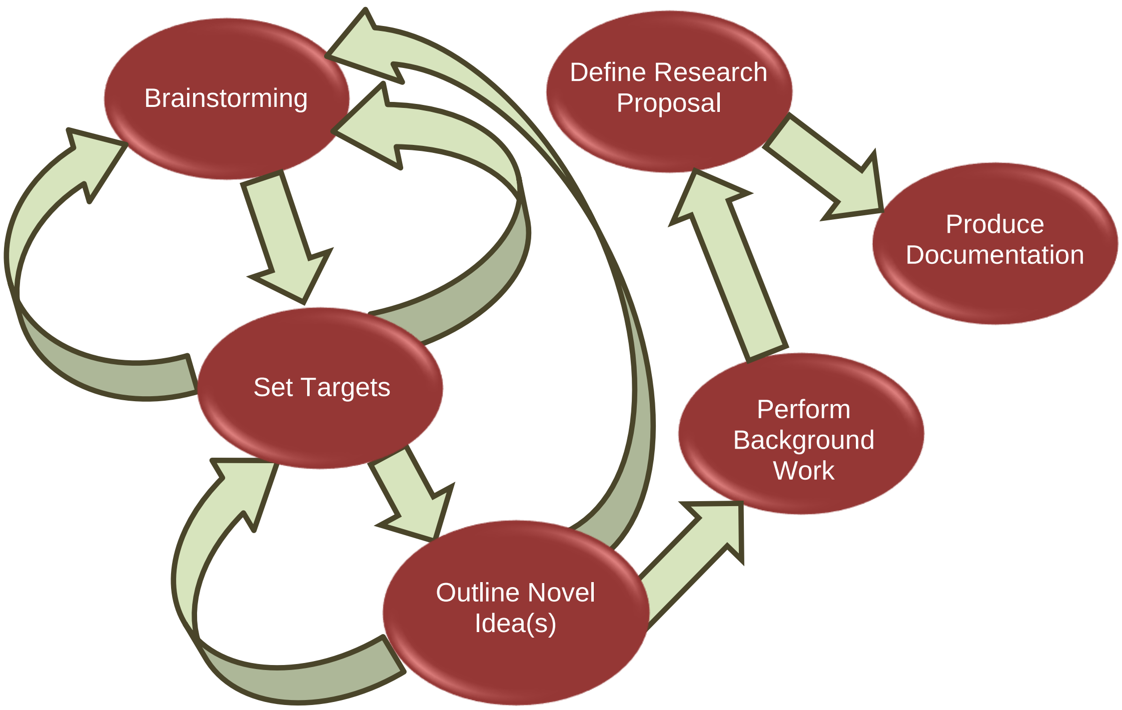

Fig.4: Staged approach for establishing and following a research roadmap

4. Conclusion

This document offered an outline of the foundational instruments enacted to drive, supervise, and incept the research trajectories interested by the project. The document provides an overview over the foundational actors used to incept the DESTINI strategy and offered a set of best practices which were utilized to mentor and tutor ESRs visiting Cyprus and collaborating with local researchers for strengthening collaboration between the partners and ultimately leading to high-value research publications.

We conclude that the best practices elicited in this document and applied for the corresponding DESTINI activities are effective in that they were able to deliver a steady flow of research within and across the units involved in the project and well beyond the prescribed expectations of the project itself.

In the future, we plan to formalize the template and offer it as open-source, for others to embrace the DESTINI model at a larger scale and beyond the scope of the project.