Mapping Our Data for Good Journey

UNIVERSITY of WASHINGTON
eScience Institute
DATA SCIENCE FOR SOCIAL GOOD

Workshop on Data for Good in Education (D4G4ED) 6/2/23

Dharma Dailey, ddailey@uw.edu

Mapping Our Data for Good Journey

Talk topics

- Mapping your D4G teaching journey (10 considerations)
- Mapping your learners journey to the Data for Good Workflow
- Opportunities for bringing good into the learning experiences we facilitate

Dharma Dailey, PhD
Research Associate, eScience Institute + UW Bothell Computing & Software Systems Part-time Faculty
Human-Centered Data Scientist for the UW Data Science for Social Good Program

Mapping Our Data for Good Journey

“Observer Notes”

In a few places, I’m adding in my own observations and perspectives. These may or may not apply to your situation.

Mapping Your Data for Good Journey

Mapping Our Data for Good Journey

10 Considerations

What kind of good are you aiming to do?
What is your Theory of Change?
What's new to you?
What is your personal contribution?
What is your relationship with those concerned?
How much time do you need?
How much time do you have?
Who are your learners?
How does the good you are aiming for relate to the subjects areas you teach?
What support can you martial?

What kind of good are you aiming to do?

Some forms of doing good
are more straight-forward
than others.

What kind of good are you aiming to do?

Data for Good problems span all of these categories and ranges

Degree of potential impact

- Intended and unintended

Novelty

- Research question, Interventi

Wickedness/Complexity

- Technological/Methodological
- Organizational
- Social

What kind of good are you aiming to do?

Novelty **Low**  **High**

- Application / Intervention
- Research question

What kind of good are you aiming to do?

Potential degree of impact **Low**  **High**

- Intended and Unintended
- Direct and Indirect

What kind of good are you aiming to do?

Wickedness/Complexity	Lower	Higher
Technological / Methodological	Tableau Dashboard with well-understood data	Novelty/Sensitivity/Scale
Organizational	One Decision Maker	Multiple less-resourced orgs
Social	High social agreement , shared understanding of the problem	• Low social agreement / controversy • Those impacted are not in decision makers / not shaping the project • Concerns vulnerable, less resourced or historically marginalized groups

What kind of good are you aiming to do?

Observer Note:

- Monitoring & Evaluation and other established, repeated, but valuable analyses are likely easier to fold into coursework than novel data and RQs
  - Aligns well with “Data basics”
  - Less prep for you if repeated
  - More possibilities for scaffolding integrative learning
- High need for smaller, less resourced organizations
  - Repeated support over time creates increased value for them

What is new to you?

- What skills, knowledge, and resources do you have on hand? What is in development?
- Buffer time for our own learning curve and building new relationships (“new x 2”)

Research question
Methods
Data
Subject matter
Collaborations
Intervention space
Stakeholders
Sponsors

What is your Theory of Change?

- **How does working with data in some manner manifest a particular social good?**
- **Who shares your theory of change? Whose version of good?**

Being able to explain your Theory of Change is key to attracting resources and collaborators, effective collaboration, getting student buy-in and setting up their contributions.

What is your Theory of Change?

- **How does working with data in some manner manifest a particular social good?**
- **Who shares your theory of change? Whose version of good?**

Observer Note:
In general, the clearer and more specific our ToC is, the more effective we'll be.
However, the reality is, we gain clarity as we gain more experience with the problem space, the intervention space, with collaborator and stakeholders .

What is your personal contribution to the good you are aiming to achieve?

“Falling in the love with the problem”

(credit to Heather Leson of the Solferino Academy of International Red Cross)

What part of the problem are you in love with?

- Methodological /technological expertise
- Subject matter knowledge
- Intervention know-how
- Impacted communities

Are you partnered with people who are in love with other aspects of the problem?

What is your relationship with those concerned?

Who trusts you and why?
Who are you accountable to and who is accountable to you?
Long-term or short-term relationship?

Observer Notes:

- Short-term can work if you can provide value to someone who “owns” the problem long-term
- I highly recommend doing a stakeholder analysis workshop with your collaborators and students for every Data for Good project

How much time do you need?

Good science is slow and incremental and the work of many hands
Good social change is slow and incremental and the work of many hands

- Yet we are “results oriented”, we want to provide “insights” and so do our students.
- We need to feel like we are making authentic contributions to stay motivated.

Observer note
Giving students the opportunity to connect directly with people who are the direct beneficiaries and the stewards for social change helps them to

- See their work in perspective
- Develop and sustain motivation
- Improves the quality of their work

How much time do you have?

Marathon runners

- Subject matter experts
- Intervention know-how experts
- Those with trusted relationships with impacted community

Methodological experts can support sprints

E.g UW DSSG teams work over the summer. Professional data scientists and advanced students with computational and statistical know work for 10 weeks with a subject matter expert who is a the project lead.

Who are your learners?

- Coaching PhD students to tackle novel scientific discoveries?
- Introducing basic analytics to first years undergraduates?

Observer note:
One of the strongest rationales for Data for Good teaching/learning is that it can attract and retain students who are underrepresented in STEM:

- Science that expresses communitarian values aids in retention of women and minoritized students to STEM (See National Academies lit review)
- Data for Good is a different kind of learning opportunity that attracts a different kind of student

How does the good you are aiming for relate to the subjects areas you teach?

Are we bringing in some subject matter and intervention know-how into a learning experience that centers upon specific computational or statistical techniques— or vice versa?

Observer note:

A Word of Caution to methods-oriented educators:

It can be tricky to start with a solution and work your way into authentic Data for Good problem solving

What support can you martial?

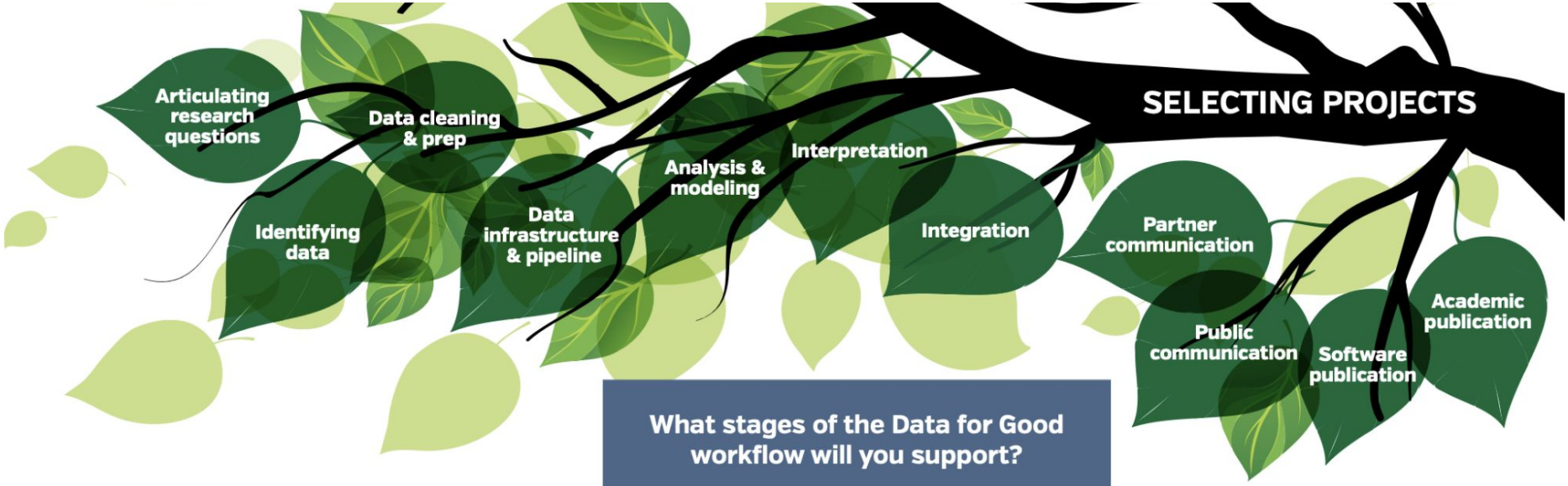
What support, resources, and collaborations can you align?
Whose expectations do you need to align?

- Community based work can have a lot of impact, but is not always valued in by the academic peers or job descriptions
- Ideally, we are resources for *all* the work involved, to make that happen might require educating different stakeholders about what is involved
- E.g. can Data for Good projects count towards your tenor package?

Mapping Your Learners Journey to the Data for Good Workflow

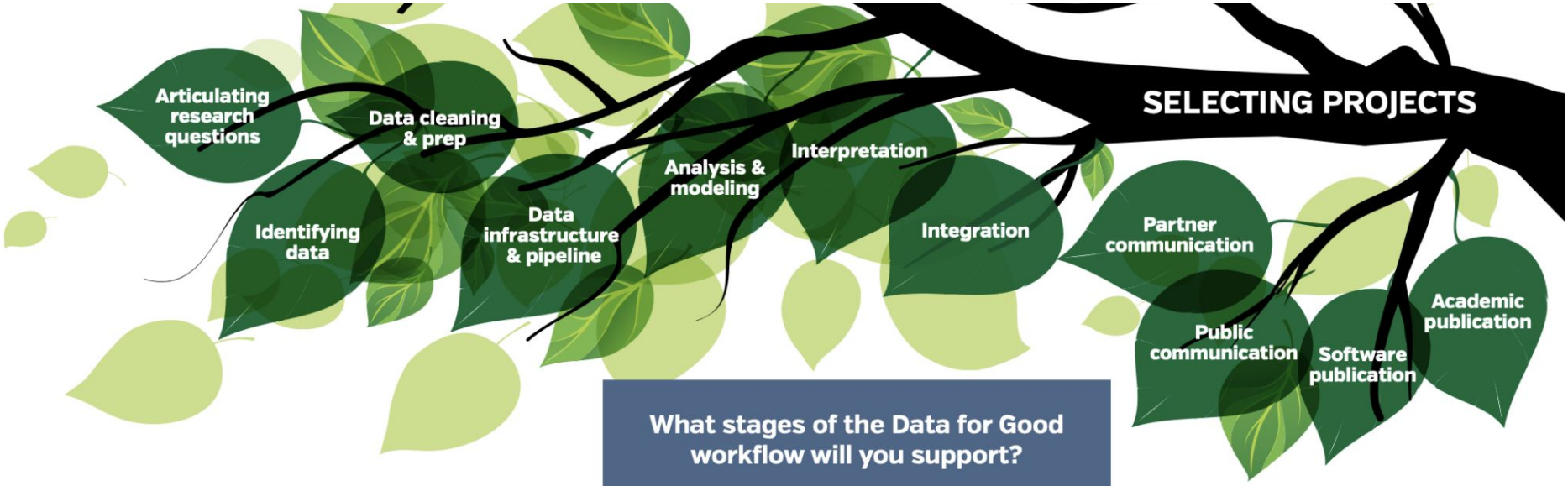
Mapping Your Learners Journey to the D4G Workflow

Choose which stages of the D4G workflow to support.



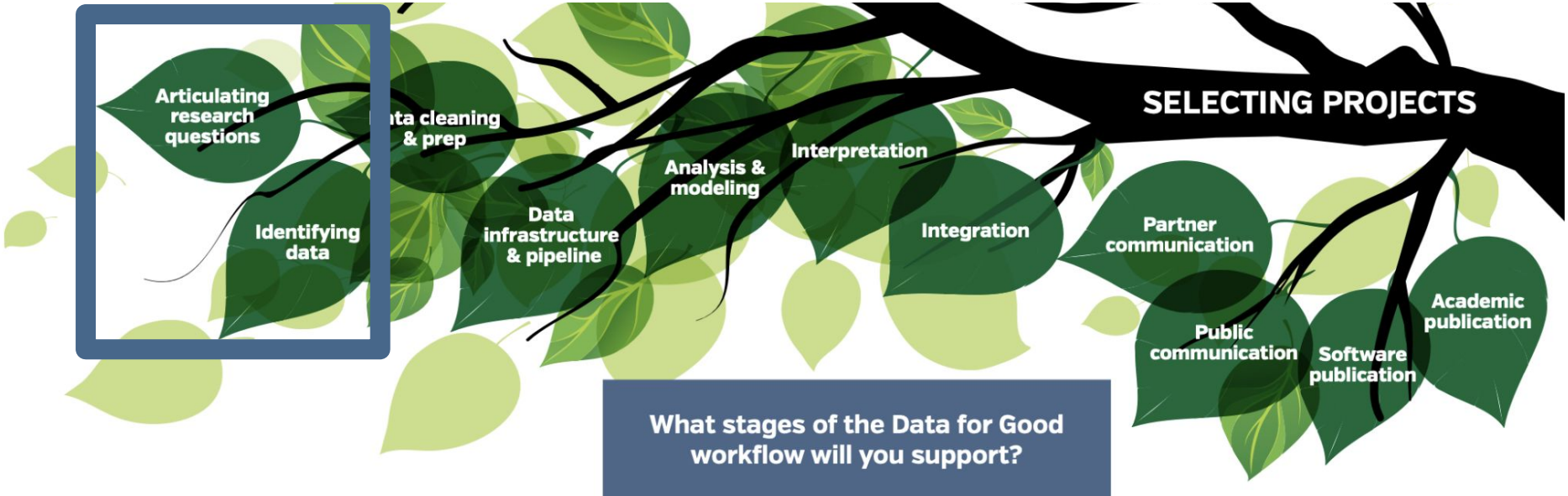
Mapping Your Learners Journey to the D4G Workflow

A D4G project can be successfully centered around any part the Data for Good workflow



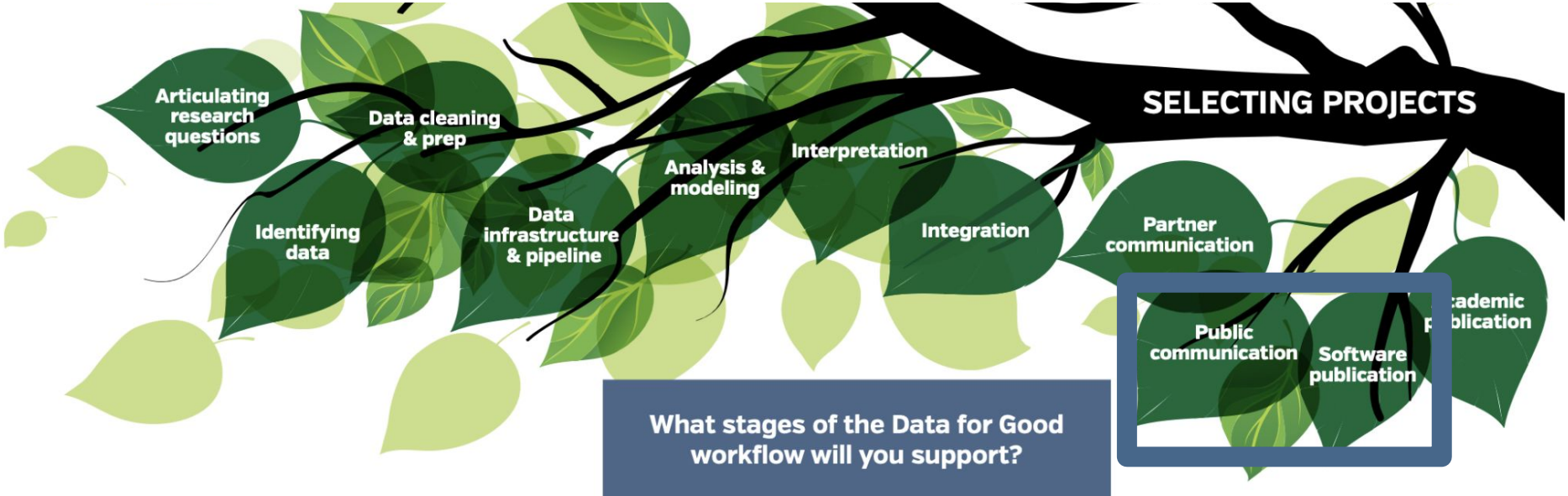
Mapping Your Learners Journey to the D4G Workflow

E.g. a D4G project may focus on helping a project partner to articulate a research question & “data discovery”



Mapping Your Learners Journey to the D4G Workflow

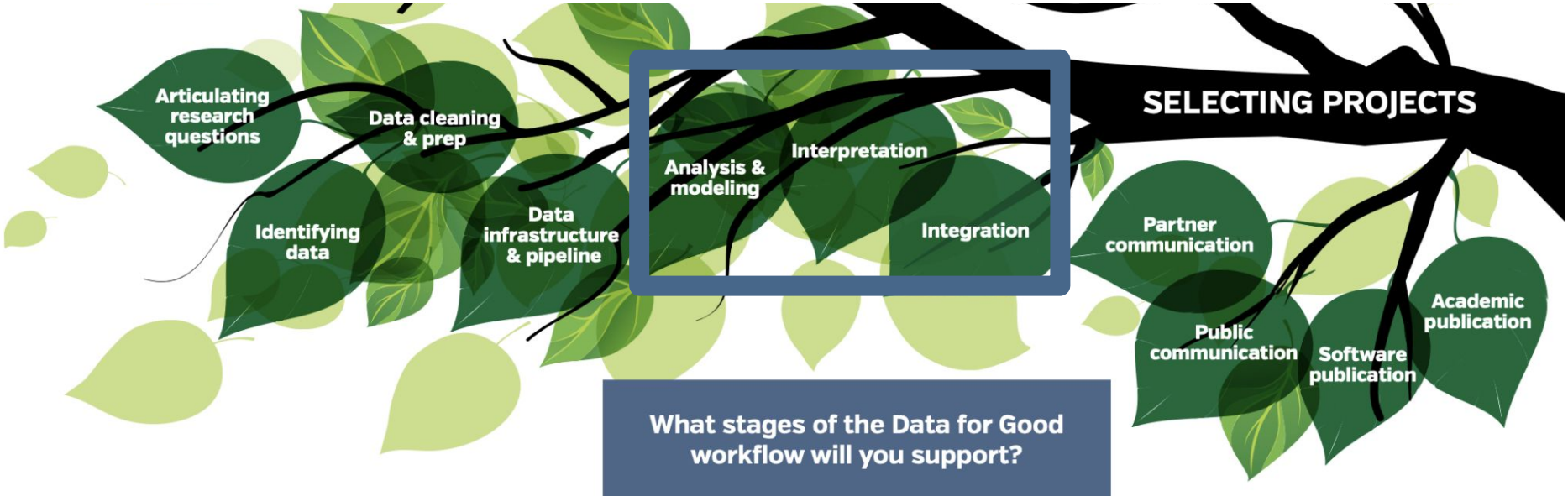
Or a D4G project may focus on making existing data science software more user friendly and more accessible



COMMON PROGRAM CHARACTERISTICS

USE DATA SCIENCE TECHNIQUES

If the goal is for students to spend most of their time modeling with shared results at the end…



Mapping Your Learners Journey to the D4G Workflow

... then prior preparation is needed

Before

- Research question
- Research plan
- Intervention plan
- Ready-to-use data

During

After

- Integrate program deliverables into an intervention

Opportunities for Bringing Good into the Learning Experiences We Facilitate

Weaving good into our the learning experiences we facilitate

Data for Good opportunities map to every level of teaching-learning

Across

- **Sectors** **(e.g.** Education + Government or Education + Community-based organizations)
- **Our Educational Institutions**
- **Our Program/Department Offerings** (e.g. projects woven through multiple courses)

Focus for

- **Course Offering, Research for Credit, Internships, Non-credit summer program, workshops**
- **Lesson Plans**

Observer note:

Integrative learning is imperative

Social and ethical challenges are baked in to data science

How do we realize the good we envision?

Phase	Social and technical considerations are entwined during each phase of a UW DSSG project:

DEFINE **PREPARE** **ANALYZE** **DELIVER**

----------> Communicate with project partners and subject matter experts ---------->

Articulate research a question

Identify data

Clean & prep data

Develop a data infrastructure & pipeline

Model & analyze

Interpret

Integrate

Public communication

Software publication

Academic publication

THE DATA FOR GOOD WORKFLOW

Phase	Social and technical considerations are entwined during each phase of a UW DSSG project:
DEFINE	A data science research question that can inform a social intervention is first and foremost shaped by subject matter experts and intervention experts. However, the plan to answer the research question must be informed by the computationally-assisted statistical techniques and software engineering that will be applied to data that is either curated or created for the planned analysis. Thus, the research question must be driven by "the social" but informed by "the technical." This might take multiple iterations with people who have complementary expertise.
PREPARE	Preparing data and building a data infrastructure for a DSSG project requires careful attention to both social and technical considerations. The intellectual work of cleaning and preparing data is often intimately tied to determining what questions can be soundly and ethically accomplished. The data infrastructure and pipeline must be designed to consider potential harms and benefits to those represented in the data, potential harms and benefits to the organizations involved, and practical matters such as who will maintain the infrastructure after the project ends.
ANALYZE	The computationally-assisted analysis and software engineering most closely associated with the moniker "data science" only become sound, relevant, and useful when they are interpreted with substantive subject matter expertise—ideally with interpretive input directly or indirectly from multiple stakeholders and peer experts.
DELIVER	Because DSSG projects are rarely standalone interventions but rather specialized work within a larger intervention space, the research question, results, and analytic processes must be explained to multiple audiences beyond the research team. The various kinds of project outputs from a DSSG (analysis, analytic processes, data infrastructure, software, etc.) must be prepared for handoff to partner organizations, analysts, and/or subject matter peers as appropriate.

Dailey, Tanweer, & Stone, 2022

Social and ethical challenges are baked in to data science

Integrative learning is imperative

Ethical considerations are braided through STEM for Good work

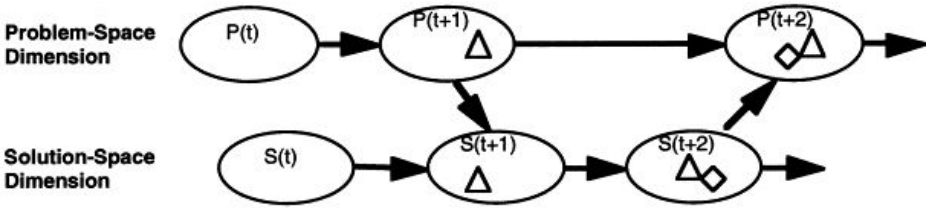
[HTML] **Creativity** in the **design process**: co-evolution of problem–solution

K Dorst, N Cross - **Design** studies, 2001 - Elsevier Paperpile

… **creativity**. From the protocol data we identify aspects of **creativity** in **design** related to the formulation of the **design** … We also apply our observations to a model of **creative design** as the co-…

☆ Save 99 Cite Cited by 3047 Related articles All 18 versions Import into BibTeX

Co-Evolution of the Problem-Solution



P(t) initial problem space
P(t+1) partial structuring of problem space

S(t) initial solution space
S(t+1) partial structuring of solution space

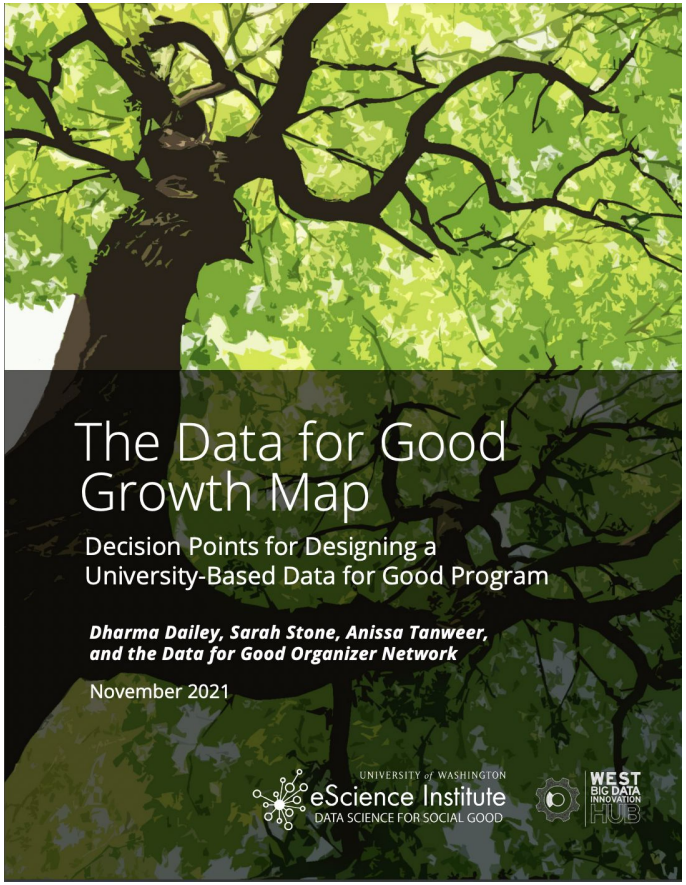
S(t+2) developed structuring of solution space
P(t+2) developed structuring of problem space

Dorst & Cross, 2001

What might we more intentionally weave together the integrative learning needed to prepare students well for what is actually takes to do good with data across the learning experiences that we facilitate?

- Computational/technological methods
- Subject matter knowledge
- Intervention know-how
- Trusted relationships with impacted community

Bonus content!!!:
D4G4ED Resources Show and Tell



The Data for Good Growth Map: Decision Points for Designing a University-Based Data for Good Program

- Focuses on decision points new programs will encounter in designing their program
- Draws on the experience of thirty-six contributors from seventeen Universities
- Focuses on summer program outside of normal courses, but may be helpful beyond that.

Link to the paper

In the shared folder for this workshop

UW DSSG Approach

UNIVERSITY of WASHINGTON eScience Institute
DATA SCIENCE FOR SOCIAL GOOD

Human Centered Design Curriculum

Activity	Detail and Rationale
Program Required	
Introduction to Data Science for Social Good* (Week 1, Day 2)	This workshop introduces the position of the program: Ethical challenges and social complexities are inherent in DSSG work. Fellows co-examine problematic case studies and discuss as a cohort. This helps fellows to name and discuss tensions in their own DSSG work as those tensions inevitably arise during the program.
Preparing for Stakeholder Engagement Workshop (Week 2)	Picking up from the Introduction to DSSG workshop, this workshop begins a dialogue on how ethical tensions and social complexities can be addressed in DSSG work. It also introduces some additional framing on the nature of DSSG work that helps fellows orient to the Data for Good workflow. Emphasis is placed on techniques that can be applied to any DSSG project: Co-inquiry, dialogue, and complementary paths of inquiry including stakeholder engagement.
Facilitated co-design, execution and debrief of a first stakeholder engagement treated as collaborative research activity (~Week 3)	The HCD mentor helps each team to design and execute an initial "stakeholder engagement." A protocol is introduced: Teams are directed to design, conduct and reflect upon the engagement as a team. This protocol helps teams to most efficiently develop a shared understanding of the project including its ethical dimensions and social complexities. It's simple enough that teams can use it on their own for the remainder of the program and beyond regardless of prior experience.
Stakeholder analysis & speculative ethics workshop* (Week 3 or 4)	After teams have become familiar with their projects, a stakeholder analysis helps teams arrive at a shared understanding for their project's potential impact and its relation to broader social context. A power analysis of which stakeholders can directly or indirectly shape the project and which stakeholders may be directly and indirectly impacted by the project clarifies the project's ethical frame, helps teams to refine project goals and better communicate the role of the DSSG project in effecting change.
Public communication via public website (Week 5, 7, 9)	A key part of HCD is translating our work for multiple audiences. All teams are required to create a public website and code repository. Cross-team writing exchanges help teams to articulate their work for a public audience of their choosing.
Optional Project-Specific	Teams typically elect to do more than the minimum required HCD activities, supported by the HCD mentor as requested. These may include experiential exercises such as "walkthroughs", site visits, interviews, user evaluations, roundtables, focus groups, surveys, and more.
Integrating HCD into project work and other program activities	HCD activities give fellows permission and tools to raise ethical and social considerations as they deem relevant throughout the program.. We deem HCD activities a success when teams are able to raise and resolve issues on their own projects and come to a shared understanding of their project's role in creating a social good. Discussions started in the HCD workshops and one-on-one team meetings are carried over in Career Panels, Talking Data Science guests., et al.

*Designed and led by UW DSSG Program Director Anissa Tanweer, other activities led by Dharma Dailey, HCD Mentor

Dailey, Tanweer, & Stone, 2022

Weaving more good into our teaching/mentoring

Integrative learning
Computational/technological methods, Subject matter knowledge, Intervention know-how, Trusted relationships with impacted community

Opportunities at every level
Across sectors: Educational Institutions, Our Program/Department Offerings
Focus for: course Offering, Research for Credit, Internships, Non-credit summer program, workshops, Lesson Plans

Heavy touch or light touch
E.g. of lightly weaving in to our methods heavy technology, computing, statistical methods courses:
Are their focusing questions on social or ethical dimensions that could be asked each week to better the ethical reasoning to methodological application?

- Better models the D4G real-world workflow
- Could be a game changer in terms of culture shift
- Normalizing ethical reasoning, integrative learning

Summing up

Summing up

Mapping Your D4G Journey

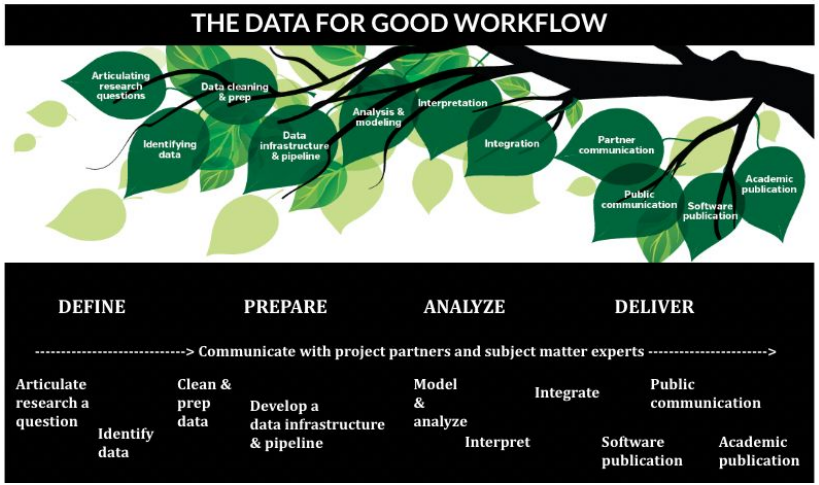
What is your Theory of Change?
What's new to you?
What is your personal contribution?
What is your relationship with those concerned?
How much time do you need?
How much time do you have?
Who are your learners?
How does the good you are aiming for relate to the subjects areas you teach?
What support can you martial?

Good News: People in your situation are innovating and making it work

,

Summing up

Mapping Your Learners Data for Good Journey



Good News: Any part of the journey is worthy of teaching/learning

Summing up

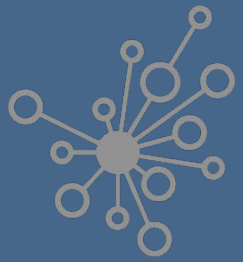
Bringing more good into what our teaching/mentoring

Integrative learning
Computational/technological methods, Subject matter knowledge, Intervention know-how, Trusted relationships with impacted community

Across Sectors
Educational Institutions, Our Program/Department Offerings
Focus for ourse Offering, Research for Credit, Internships, Non-credit summer program, workshops, Lesson Plans

Heavy touch or light touch

Good News: We are spending the rest of our weekend exchanging ideas on how to do so



UNIVERSITY of WASHINGTON
eScience Institute
DATA SCIENCE FOR SOCIAL GOOD

Thank you

ddailey@uw.edu