APPENDIX A: BLUEPRINT WORKSHEETS

To supplement The Open Data Policy Lab's Blueprint to Unlock New Data Commons for AI, this document provides a series of worksheets that organizations that have or steward data can use when creating a data commons. This includes, for instance, knowledge institutions (such as libraries, universities, research centers) and others that hold unique and high-quality information assets (such as governments, non-profits).

These worksheets correspond with the main decision points of each Module within the Blueprint: Mapping the Supply and Demand, Unlocking Participatory Governance, Building the Commons, and Assessing and Iterating. They provide checklists of elements to consider when setting up a data commons.

MODULE A: MAPPING THE SUPPLY AND DEMAND

Module A seeks to understand why AI systems need data, what data can be made available to train, adapt, or augment AI, and what a viable data commons prototype might look like that incorporates stakeholder needs and values.

A1: MAP THE DEMAND FOR DATA

Considerations	Notes
☐ Who holds and is looking for data?	

☐ Why are data scientists, researchers, and public-interest AI developers looking for data?	
☐ What public-interest AI applications are needed most?	
☐ What questions are data scientists, researchers, and public-interest AI developers looking to address?	

A2: CREATE AN INVENTORY OF THE DATA SUPPLY

Considerations	Notes
☐ What data assets can be made available by whom and to whom?	
☐ From a technical and logistical perspective, what is required to make this data available in the commons?	
☐ Is it feasible to make this data available?	

☐ Does a data commons for these data assets already exist? If so, what's missing?	

A3: UNDERSTAND STAKEHOLDER NEEDS AND PREFERENCES

Considerations	Notes
☐ What are stakeholders' values, beliefs, expectations and attitudes around making datasets a shared resource? Why do they feel this way?	
☐ Who are stakeholders comfortable with engaging with? Who are they not comfortable engaging with?	

☐ What kinds of data are stakeholders comfortable with providing access and to whom?	
☐ What data should not be made available in the commons	

☐ How will stakeholders benefit from the commons–beyond monetary benefits?
    ☐ Gain early access to the AI tools developed?
    ☐ Receive a free membership to AI tools that use their data?
    ☐ Receive insights on data about or for them?
    ☐ Gain access to the infrastructure to replicate AI tools for other contexts?	

A4: ASSESS THE PRACTICALITY OF THE COMMONS

Considerations	Notes

☐ What might a data commons concept look like?
   ☐ Governance: How will the data commons be managed? What processes, standards and regulations are needed? Who will manage them?
   ☐ Finances: How will the data commons be funded? What financial resources are needed to build and maintain the data commons? Who needs to supply these resources?
   ☐ Infrastructure: What technologies are needed to build the data commons itself? What technological infrastructure is needed to ensure there is secure access to data? Will this infrastructure be interoperable with other systems?
   ☐ Expertise: What kind of expertise is needed to set up and maintain the data commons? What human resources are needed?	

☐ What are the goals of the data commons and how will you achieve them?	

☐ What is the theory of change?	
☐ How will you measure success?	

MODULE B: UNLOCKING PARTICIPATORY GOVERNANCE

Module B focuses on co-designing key aspects of the data commons with key stakeholders and documenting these aspects within a formal agreement.

B1: CO-DESIGN HOW THE DATA COMMONS WILL BE GOVERNED

Considerations	Notes
☐ How can the specifics of the commons be co-developed with stakeholders?	

☐ How will data users and AI developers access the data in a way that is equitable? Should this involve tiered access, federated learning, commercial vs. non commercial access, sandboxes, or trustee model?	
☐ What templated agreements are needed for data supply, data access, and data use?	
☐ How will the data commons be funded? What funding models can be sustained over time (e.g. patronage, micropayments, membership fees, license and re-use fees, public-private partnerships, crowdfunding, or endowments)?	
☐ What technical infrastructure is needed (e.g. cloud-based infrastructure, generative AI applications, automated quality assurance)?	

☐ What are the accountability mechanisms? What happens if someone violates the rules of the commons?	
☐ How long will this operate (e.g. for a set period, indefinitely, until a situation passes)?	

B2: FORMALIZE THE ELEMENTS OF THE DATA COMMONS

Considerations	Notes
☐ Is there consensus?	
☐ How can we put the aspects of the data commons into practice?	
☐ Who needs to document these aspects?	

Considerations	
☐ What should be included in the data sharing, data use, and data access agreements?	

MODULE C. BUILDING THE COMMONS

Module C aims to establish the data commons from a practical perspective and ensure all stakeholders are incentivized to implement it.

C1: IN PRACTICAL TERMS, BUILD THE DATA COMMONS WITH PARTNERS AND STAKEHOLDERS

Considerations	Notes
☐ What are the roles and responsibilities of the commons? Who is involved and how?	

☐ What is needed to make the data available?	
☐ What technical infrastructure is being harnessed?	
☐ How will the data be made available?	
☐ Who should know about and participate in the data commons?	

C2: MANAGE HOW THE COMMONS OPERATES

Considerations	Notes
☐ Are the data users and contributors complying with the rules of the commons?	
☐ Are stakeholders receiving the benefits of the commons?	
☐ What does it mean to be ethical within the commons?	

☐ Are stakeholders incentivized to participate in the commons? What other incentives are needed?	

MODULE D: ASSESSING AND ITERATING

Module D involves evaluating how the commons is working and iterating as needed.

D1: EVALUATION AND LEARNING

Considerations	Notes
☐ How often should partners be engaged?	
☐ How is the data commons performing in terms of its KPIs?	

☐ Has the data on the commons been sufficient to develop public-interest AI applications? What needs to be improved?	
☐ What problems are emerging?	

D2: ITERATION

Considerations	Notes
☐ What changes need to be made in order for the data commons to be sustainable?	