HDR DSC: The Metropolitan Chicago Data-Science Corps (MCDC): Learning from Data to Support Communities

AIMS:

Facilitate the transformation of data into knowledge that is inspiring for and actionable by a community organization.

Contribute to an agile, representative, and inclusive data science workforce via designing and implementing cross-disciplinary training, mentoring, and opportunities for diverse student groups to acquire broad and transferable analytical skills.

Build a community of educators, researchers, and students that can work with communities to discover and disseminate best practices in data science.
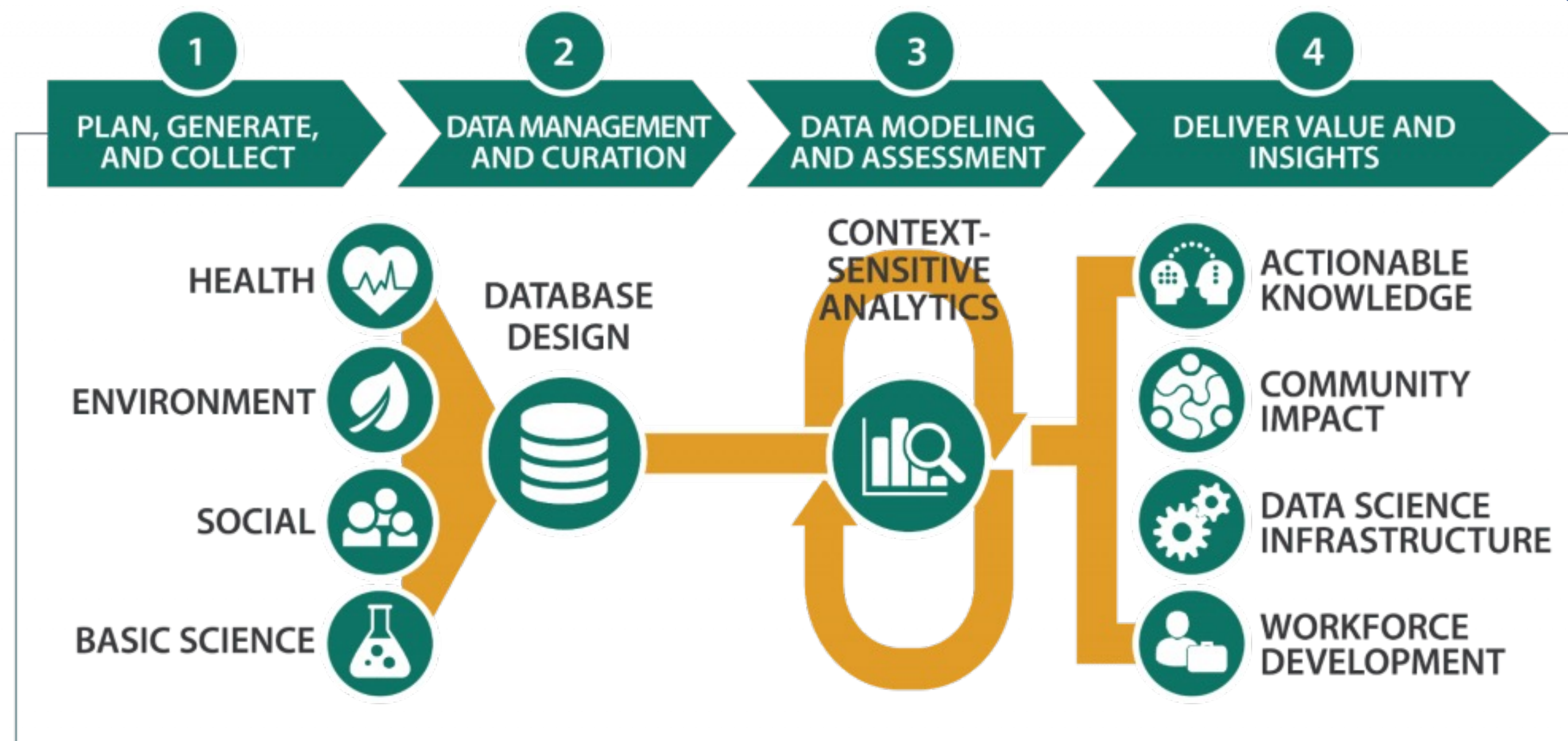
Facilitate access to data, real-world data science projects, and data science training materials for the common good.

The MCDC focuses on the intersection of data science with three domain-science areas: social sciences, environmental sciences, and health sciences.

This intersection is supported by fundamental science and connected through shared analytical and scalable methodologies, cross-fertilization, innovation, and a focus on equity.

The MCDC curriculum focuses on nine areas of data acumen that are interwoven in a Data Refinery model, depicted here. Data refining begins with the process of planning, generating and collecting data, and ends with the delivery of insights and knowledge extracted from the data and of value to a community partner.

Educational Structure

1. Students can take **multiple pathways** through a wide range of foundational data science courses, addressing data acumen 1-4.
2. Student participate in a **practicum**, addressing data acumen 5-9. Students work on a real-world data science problem presented by a community partner, while the focus remain on academic and professional development of the students.
3. Student may participate in an 8-week summer internship (**DAU**) in partnership with community organization and a faculty mentor. Students enter in a more deliverable-oriented relationship with the community partner.

Data Acumen

1. *Data management and curation*
   Effectively and efficiently import, clean,. Transform, and combine data (data preparation and wrangling). Handle data of various formats.
2. *Data description and visualization*
   Select and construct visual representations of data for exploration and communication.
3. *Data modeling and assessment*
   Fit competing models to data. Regression, regularization, non-linear modeling, resampling, SVM, classification, cross-validation.
4. *Workflow and reproducibility*
   Analytic code that follows standards, is documented, and independently reproducible.
5. *Communication and team work*
   Effective oral and written communication with team members, community partners, mentors, diverse audiences.
6. *Domain-specific considerations in health, environmental, or social sciences.*
   Deeper knowledge of a domain science and its data.
7. *Ethical problem solving*
   Appreciate multiple perspectives, esp. marginalized communities. Promote transparency, equity, community partnership.
8. *Critical thinking.*
   Question applicability of analytical methods, engage in meaningful thought experiments, understand limitations of outcomes, avoid pitfalls.
9. *Creativity and entrepreneurship*
   Synthesize, enhance beyond the sum of the parts.