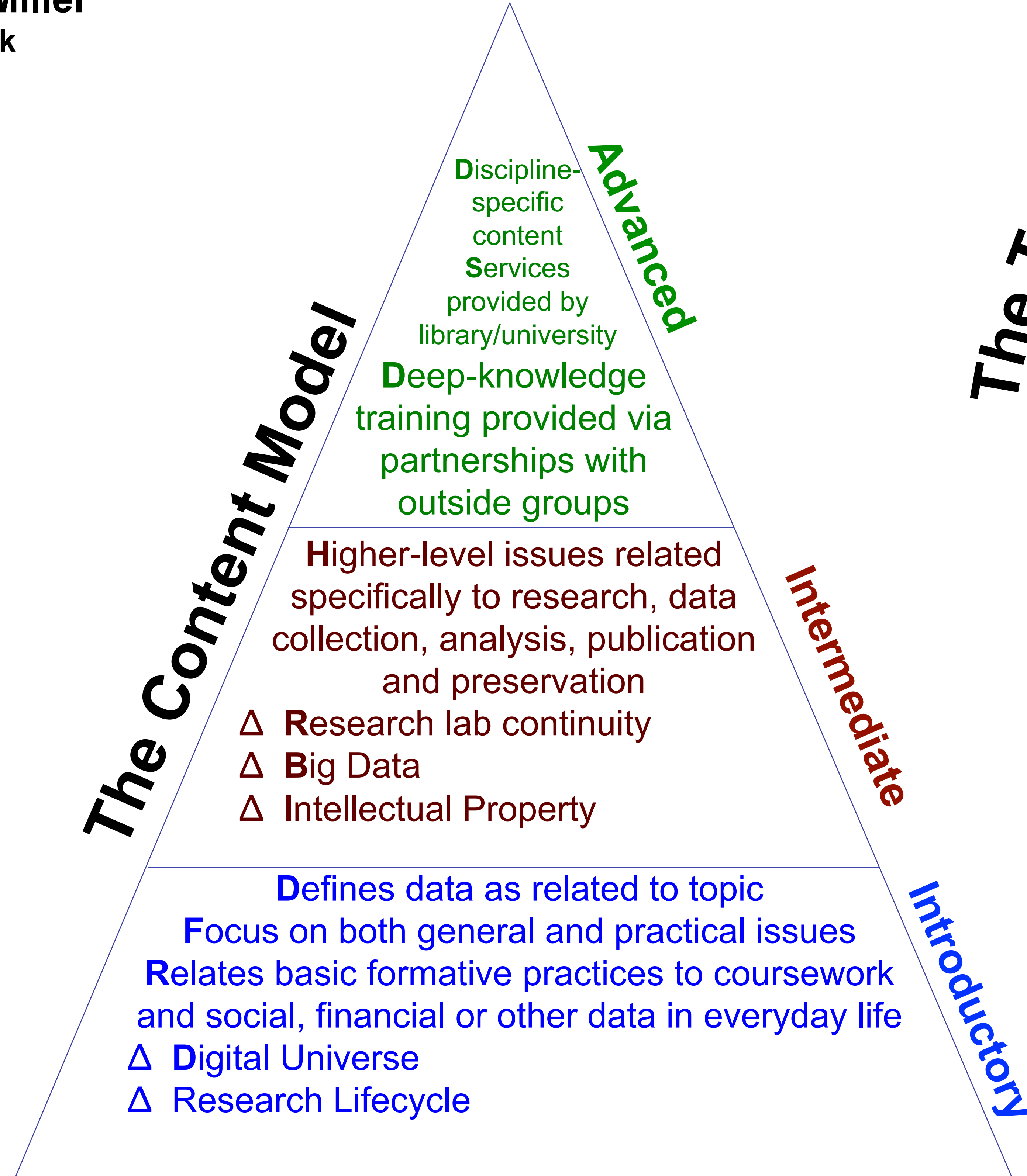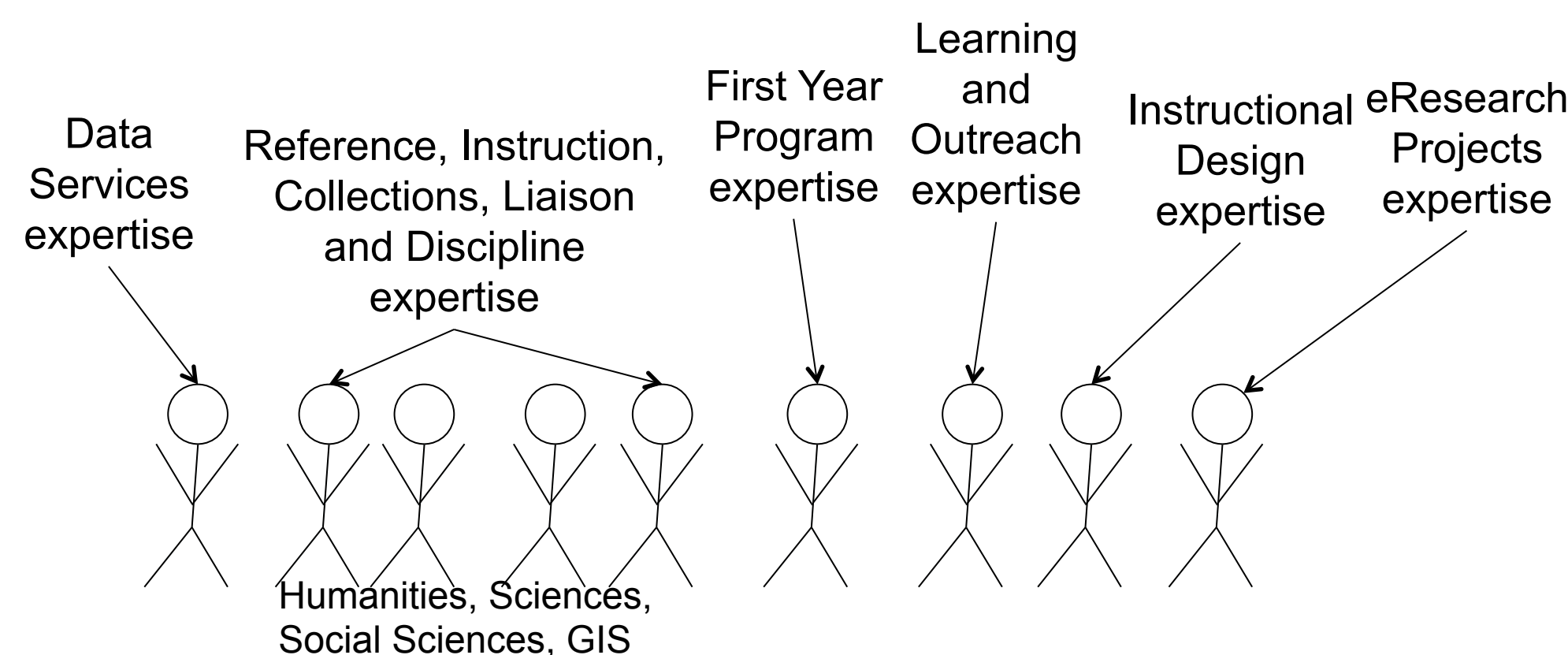# The Data Literacy Advisory Team at Virginia Tech:
Developing a Content Model for Data Literacy Instruction

Andi Ogier, Edward Lener and Rebecca K. Miller
with special thanks to Carolyn Meier and Bruce Pencek

Creation of DLAT

In Spring 2013, the University Libraries at Virginia Tech formed a Data Literacy Advisory Team (DLAT), an interdisciplinary library team. Tasked with developing an institutional framework for delivering research and data-related content to students, staff and faculty, DLAT began by developing a skill-focused content model that could be easily modified to engage audiences with different levels of experience. The skill-focused model allows instructors to select curated content from the modular framework and gives them the opportunity to modify and refine it according to their chosen learning outcomes and instructional needs.

Skill Composition of DLAT

Data Services expertise

Reference, Instruction, Collections, Liaison and Discipline expertise

First Year Program expertise

Learning and Outreach expertise

Instructional Design expertise

eResearch Projects expertise

Humanities, Sciences, Social Sciences, GIS

The Content Model

Advanced

Discipline-specific content **S**ervices provided by library/university
Deep-knowledge training provided via partnerships with outside groups

Intermediate

Higher-level issues related specifically to research, data collection, analysis, publication and preservation
Δ **R**esearch lab continuity
Δ **B**ig Data
Δ **I**ntellectual Property

Introductory

Defines data as related to topic
Focus on both general and practical issues
Relates basic formative practices to coursework and social, financial or other data in everyday life
Δ **D**igital Universe
Δ **R**esearch Lifecycle

The Topic Model

Data Management and Organization
Δ Defines data, big data, data management
Δ Data management best practices
→ coursework and management of social and financial data
Δ Emphasis on organization, citation, confidentiality
Δ Focus on research and lab ethics, data documentation, and version control
Δ Lab Notebooks and the importance of documentation and data management to Lab continuity
Δ Funding Agency Info
Δ Local DMP Services

Data Conversion and Interoperability
Δ File formats, including examples of specific disciplinary format standards
Δ Non-proprietary format standards
Δ Master copies and redundant storage
Δ Benefits of migration to downstream use
Δ When and why migration is important
Δ Benefits and risks of emulation
Δ Planned and unplanned obsolescence

Data Sharing and Acquisition
Δ Importance of data management practices to sharing data with colleagues, PIs, advisors
Δ Introduces repositories
Δ Identification of trusted data sources and trusted repositories
Δ Case for data sharing
Δ OSTP Mandate and current federal public access policies
Δ Examples from RetractionWatch

Ethics and Responsibility
Δ Ethical issues in research and scholarship
Δ Research Roles
Δ Citation (with examples from RetractionWatch)
Δ Introduction to Intellectual Property and Copyright
Δ Lab Ethics
Δ Data Ownership
Δ Confidentiality and Privacy
Δ Ethical Responsibilities in research
Δ IP and Copyright
Δ Open Access
Δ Open Data
Δ Fixity
Δ Generating Checksums
Δ Open Access Publishing
Δ Local IR

Metadata and Quality Assurance
Δ Importance of Metadata in data sharing
Δ Using metadata to analyze datasets
Δ Structure and purpose of metadata schema
Δ Metadata as documentation, annotation, and description
Δ Organizing data
Δ Table headers
Δ Naming conventions
Δ Metadata and lab continuity
Δ Disciplinary schema
Δ Open Data

Preservation and Archiving
Δ Introduction to Archives
Δ Issues in preservation of digital objects
Δ Benefits of Preservation
Δ Potential for future use/reuse of data
Δ Preservation as a part of Data Management
Δ Preservation Strategies and Technologies
Δ Reproducibility
Δ Costs of Preservation
Δ Local Repositories

Data Curation
Δ Data Curation Timeline – relates to Data Lifecycle
Δ Complexity of Research Data
Δ Data Curation Planning
Δ Lifecycle of Research Data
Δ Potential for Reuse beyond original purpose
Δ Potential Value-Adds for Data
Δ Data Curation, Validity and Reproducibility
Δ Visualization
Δ Data Mining
Δ Machine Learning

Data Science
Δ Data Science as
Δ Programming +
Δ Statistics +
Δ Domain Knowledge +
Δ Storytelling (Visualization)
Δ Data Science Resources
Δ Data Science Examples
Δ Amazon
Δ Twitter
Δ Big Data + Data Science
Δ Telling stories with data

?