**ABSTRACT:**

In response to the current data-intensive research environment and the necessity of a professional data workforce, iSchools are building new programs and enhancing existing programs to meet workforce demands in data curation, data management, and data science [1-3]. To understand the state of education in the field, we studied current programs and courses offered at iSchools and other schools of Library and Information Science. Here we present an overview of the methods and results. Courses are divided into four categories: data-centric, digital inclusive, digital, and traditional. The analysis reveals trends in LIS education for data professionals and identifies particular areas of expertise and gaps in LIS education for data professionals.

**Course Category Distribution**

Traditional: 54.6%

Digital: 27.6%

Data Inclusive: 18.8%

Annotated:

**Methods:**

- Courses and programs identified by searching online course catalogs. Searches limited to courses in library or information schools.
- Either the course name or description had to include a keyword or keyword combination associated with data curation, data science or data management.
- Data Curation broadly defined as the active and ongoing management of data through its lifecycle of interest and usefulness to scholarship, science, and education. Data curation activities enable data discovery and retrieval, maintain data quality, add value, and provide for re-use over time. This field also includes authentication, data standards, archiving, collection and management, preservation, retrieval, knowledge representation, and policy as it affects data.
- To further clean and validate the dataset, course descriptions were viewed in context and individuals at each institution were contacted (54.7% return).
- The dataset contained 476 courses in 158 programs at 55 institutions.
- Course and program descriptions were coded separately using AtlasTi by selecting every descriptive word or phrase and then grouping codes into families associated with data curation or data science as found in the literature.

**Final Search Terms:**

- Archiving (digital or data)
- Authentication (data)
- Conservation (digital or data)
- Curation (digital or data)
- Cyberinfrastructure
- Data access
- Data collection
- Data discovery
- Data mining
- Data provenance
- Data quality
- Data retrieval
- Data standards (non CS)
- Digital library
- Digitalization
- Ontology
- Policies (digital, data or info.)
- Preservation (digital or data)
- Representation (data, info., or knowledge)
- Retrieval (digital, data, or info.)
- Semantic web
- Systems analysis

Keywords derived from definition, current literature on data curation, and by consulting the Matrix of Digital Curation Knowledge and Competencies [4]. They evolved as new and relevant terms were identified.

- There were over 800 different terms in 13 families of concepts.
- The most common terms were Metadata, Preservation, Retrieval, Archives, Management, Organization, Indexing, Human Computer Interaction, and Digital Library. Each had sub entries such as Management (Asset, Digital, Data, Electronic, Information, Knowledge, Records, Systems, & Theory)
- 172 course descriptions specifically mention the word “data” in some context; however some were research methods courses.
- Data Mining was the most common usage of data in data centric or data inclusive courses. Representation & Modeling (which included Metadata codes) and Management (which included Data Management) were the next most common occurrences of data in course descriptions.
- Management was the most common concept represented in courses and by institution followed by Representation & Modeling; Information systems and Administration; Discovery, Access, & Use, coinciding with search term representation in course descriptions.
- No institution represented every concept area. On average, 6.0 concepts were represented per institution, with no institution representing more than 11 of the 13 code families.
- Despite search criteria, most courses were still traditional courses that covered some form of data curation topic.

**CONCLUSIONS:**

iSchools are making important progress on curriculum for educating the data workforce, but there is high dependency on existing digital library curriculum and limited new curriculum specific to “research data” expertise.

- 11 institutions offer 5 programs specifically focused on data and 12 more covering aspects of data; 15 other institutions have programs with a pronounced emphasis on digital content, but not necessarily data.
- There is wide variability in the terminology used to describe courses and concepts.
- Most data-centric courses appear new, with entry-level course numbers, and most coverage of data issues and expertise appears to be through revision of existing courses.
- Existing digital courses are contributing the most to data-oriented content, covering areas such as representation & modeling and archiving / preservation, but traditional courses make up the majority of courses overall.
- Data and digital categorizations apply to both program and course level attributes.
- Schools lacking well-defined “digital” or “data” curriculum offer access to some courses of value to students wishing to develop expertise as data professionals.

Further investment in data centric courses and programs will be essential to support contemporary science and research.

**Program Descriptions Highlighting Data or Digital Aspects**

- Data Centric programs were named Information Architecture, Informatics, Data Curation, Knowledge & Data Discovery, and e-Science specifically. Data Inclusive programs covered the areas of digital curation; informatics; digital libraries; escience; information architecture; information, records, content, or knowledge management; and archives & preservation.
- Data Centric programs emphasized data discovery, collection, indexing, access, retrieval, representation, sharing, mining, analysis, standards, modeling, policy, management, metrics, preservation, and archiving in their descriptions and course representation.
- No trends were found regarding whether courses were required or recommended in programs.

- 26.2% of courses analyzed were available online & 50.8% of those were exclusively online. Online only courses tended to be newer digital or data-centric courses.