



What is data literacy and why are librarians the best people to support it?

UKSG eNews 558

Joanne Fitzpatrick, Research Data Manager, Lancaster University

In today's digital landscape, data reigns supreme, wielding unparalleled influence over our lives throughout our day, informing our perspectives, shaping our choices, and driving decision making. In a world awash with big data, the need for Data Literacy support has never been more critical.

Every organisation, every professional and every individual now has need to develop a basic understanding of data and how to 'read' it, which is captured well with the phrase 'Data Literacy'. Data Literacy, defined as the ability to understand and communicate with data, has transcended its niche status as being the preserve of data analysts and data scientists, and has become a vital skill for all.

This editorial will focus on the specific competencies that encompass Data Literacy, with the aim of showcasing how relevant this is to the role of the librarian, and can be thought of in a similar way to Information Literacy. Just as Information Literacy empowers individuals to critically evaluate and use information, Data Literacy equips them to navigate the data-saturated world intelligently. These competencies align closely with the librarian's traditional role as stewards of information, making librarians well-suited to guide individuals and communities through the labyrinth of available data, just as they do now for collections.

Gathering

The first area can be thought of as 'gathering' activities, which not only includes a focus on re-use via an assessment of the data landscape and what is accessible within it, but also can include low level original data collection, for example, through Freedom of Information requests. Gathering is much more than searching however, and involves examining datasets to locate relevant data within them, extracting data that will support the original question via migration (which could be as simple as copy and paste, and as complicated as using an API) and using filtering and slicing to refine available data.

Formatting

Librarians concerned with digital preservation are already experts in file types which is one aspect of formatting, however formatting within datasets is considered here as well. This involves a skilled consideration of appropriate formatting, and thinking of the difference between how numbers could showcase currency, a date, or a phone number gives you an indication of what this area includes. Metadata standards can inform formatting as well, through aspects such as controlled vocabularies or standardised structures.

Cleaning

Cleaning data involves two steps that are equally important: finding faults, and correction. Searching for faults can be tough when ‘you don’t know what you don’t know’, but cross checking where possible and using simple functions within data manipulation software in a well thought through way can make big improvements. Fault finding can include searching for spelling mistakes, unexpectedly large or small values, blanks, duplicates and incorrectly formatted data. It is the next step after that to take action to correct this where possible, which might involve automated or manual methods, or removing parts of the data as unusable.

Linking

Blending and linking data is where you can really add value, and all you need to do this well is a common field in each of the datasets you are working with, often called the ‘primary key’. If this is not available and the data is unstructured or not standardised, adding datasets together can still provide extra insight, for example, the same data across different time periods or regions linked at a common metric is a simple yet powerful way to blend data.

Basic Statistics

The statistical analysis aspects of Data Literacy can be where individuals feel most intimidated, but you can see how far we have come already before encountering statistics, and in order to be data literate, only a basic understanding is needed in this area in order to have big impact. You may only ever require simple calculations, such as grand totals, averages, or percentages but what is more important, is whether that is an appropriate and relevant statistical analysis to use, and that is where the skill lies.

Pattern Finding

Finding the insight, or the story, in the data is what positions data as being so important to the world today. This can be achieved by a wide range of observations: summarising data, finding trends, mapping changes, uncovering correlations, comparing values, and finding highest and lowest values. The key is to do this in a meaningful way that meets the needs of the original reasons for seeking data to begin with, and you can see how relevant librarian’s skills are to this area.

Legislation and Policy

Working with data involves compliance with a range of legislations and security policies that can again seem complex and intimidating, but overlaps with many of librarian’s existing skills. Legislation might encompass GDPR, Intellectual Property and licensing, all areas that librarians currently operate within, and security considerations can be simple housekeeping

procedures such as managing cloud access, password protecting files and using recommended organisational systems.

Visualising

Closely linked with pattern finding and communicating, visualising data can be as simple as preparing a relevant graph or chart, but can also include more advanced visualisations and infographics, and could even be 'live' changing visualisations that can be controlled by the consumer of the final data output.

Communicating

Finally, a full consideration of communication is essential when considering data. What details to include, such as statistical manipulations, gathering methodologies or full overviews of patterns, can be decided upon depending on the audience, where certain end users would prefer to have these omitted and just the final insights to be communicated through visualisations and story telling. Accessibility and inclusion is another addition to this area, and elements such as accessible documents, representation and bias are where librarians are well placed to enhance skills.

In conclusion, this editorial seeks to demystify the world of data and build the confidence that this is definitely something librarians can work with themselves, and they can also support others to do the same. Data Literacy holds the context in which data appears as equally important as the statistical analysis and coding that more advanced data work can also involve. In addition to the skills described above, this context includes a deep understanding of the data's source, its reliability, and its ethical implications. Furthermore, Data Literacy extends beyond just the technical aspects, encompassing critical thinking and the ability to communicate data-driven insights effectively. Librarians, with their expertise in information management, research, and education, are well situated to play a pivotal role in fostering Data Literacy within their communities. By bridging the gap between the technical and contextual facets of data, librarians can empower individuals and organisations to navigate today's dense data landscape, fostering a more informed and data-literate society.