

**DATA CURATION  
(DSC3141)**

**Time Allotted : 2½ hrs**

**Full Marks : 60**

*Figures out of the right margin indicate full marks.*

*Candidates are required to answer Group A and any 4 (four) from Group B to E, taking one from each group.*

*Candidates are required to give answer in their own words as far as practicable.*

**Group – A**

1. Answer any twelve:

**12 × 1 = 12**

*Choose the correct alternative for the following*

- (i) Point out the correct statement  
(a) Raw data is original source of data  
(b) Preprocessed data is original source of data  
(c) Raw data is the data obtained after processing steps  
(d) None of the mentioned
- (ii) Among the following options choose which one of the following focuses on the discovery of unknown properties on the data  
(a) Big data  
(b) Data mining  
(c) Machine learning  
(d) Data wrangling
- (iii) What data type is generated by imitating the operation of a real-world process or system over time using computer test models?  
(a) Observational data.  
(b) Experimental data.  
(c) Simulation data.  
(d) Derived / compiled data.
- (iv) Which of the following is not a use case of data catalog.  
(a) Efficient data curation  
(b) Unifying all data context  
(c) A data warehouses  
(d) Cost optimization
- (v) Which of these are the advantages for using a data management service (DMS) (select only one)?  
(a) data redundancy and inconsistency is reduced  
(b) data movement is managed  
(c) data governance is promoted  
(d) all of these are correct
- (vi) Which of the following is NOT listed as a type of content archive managers must deal with?  
(a) Audio/music files  
(b) Social media content  
(c) Virtual Reality environments  
(d) Images and videos

- (vii) One of the challenges that libraries face regarding data curation is:  
 (a) The need to integrate more data sources that previously  
 (b) Users are getting less interested  
 (c) Data curation is less well understood  
 (d) None of the answers is correct
- (viii) what is arXiv?  
 (a) Peer reviewed pre-prints for the field of physics  
 (b) Repository of electronic preprints of scientific papers in the field of mathematics, physics, quantitative biology and others.  
 (c) A group of archives spanning many different scientific disciplines.  
 (d) Preprint repository for the biological sciences hosted by Cold Spring Harbour Laboratory
- (ix) Which of the following is the common goal of statistical modelling?  
 (a) Inference (b) Summarizing  
 (c) Subsetting (d) None of the above
- (x) What does the OAIS reference model primarily address?  
 (a) Digital content monitoring procedures  
 (b) Analyzing digital preservation challenges and recommending standards  
 (c) Establishing financial protocols for digital repositories  
 (d) Creating guidelines for digital content creators

*Fill in the blanks with the correct word*

- (xi) \_\_\_\_\_ is the process of moving data that is no longer actively used to a separate storage device for long-term retention.
- (xii) \_\_\_\_\_ is a formal document that outlines how data are to be handled both during a research project, and after the project is completed.
- (xiii) What does 'OAIS' stands for \_\_\_\_\_.
- (xiv) \_\_\_\_\_ is a means of managing data that makes it more useful for users engaging in data discovery and analysis.
- (xv) \_\_\_\_\_ is the process of importing large, assorted data files from multiple sources into a single, cloud-based storage medium.

### **Group - B**

2. (a) *“Data curation is the active and on-going management of data through its lifecycle of interest and usefulness; Curation activities enable data discovery and retrieval, maintain quality, add value, and provide for re-use over time”* -- Analyze the data dynamics in which data curation is inserted, investigate future requirements and emerging trends for data curation. [[CO3] (Understand/IOCQ)]
- (b) Differentiate between Data Curation, Digital Curation and Digital Stewardship. [[CO3, CO4] (Understand/IOCQ)]

**6 + 6 = 12**

3. (a) What is Metadata Management? What is the Difference Between a Data Catalog and Metadata Management? [[C03](Understand/IOCQ)]
- (b) What is the challenge in data sharing and what measure do you take before sharing the data in public. [[C02] (Remember/LOCQ)]
- (c) What are the benefits of managing research data? [[C02]Remember/LOCQ)]
- (2 + 3) + (2 + 2) + 3 = 12**

### Group - C

4. (a) Explain how the data life cycle represents the various stages of data curation process. [[C03](Understand/IOCQ)]
- (b) “Data curation can be a resource-intensive and complex task, which can easily exceed the capacity of a single individual.” – Give proper reason to justify the statement. [[C02, C04] (Understand/LOCQ)]
- (c) Distinguish the role of data manager and data curator. [[C05](Identify/IOCQ)]
- 6 + 3 + 3 = 12**
5. (a) What is digital curation and why is it important? [[C02](Understand/LOCQ)]
- (b) What are the core components of a digital curation strategy? [[C03] (Understand/IOCQ)]
- (c) How does digital curation differ from digital preservation? [[C04](Understand/LOCQ)]
- 6 + 3 + 3 = 12**

### Group - D

6. (a) What is metadata and what are its primary functions? [[C01](Define/LOCQ)]
- (b) How do you create and manage metadata standards for different types of data? [[C04](Understand/LOCQ)]
- (c) What are some best practices for ensuring metadata quality and consistency? [[C01,C02](Understand/LOCQ)]
- 3 + 3 + 6 = 12**
7. (a) What are the best practices for automating the data ingestion process? [[C01](Define/LOCQ)]
- (b) How do you handle data quality issues during ingestion? [[C04](Understand/LOCQ)]
- (c) How can you ensure that data ingestion processes are scalable and efficient? [[C01,C02](Understand/LOCQ)]
- 3 + 3 + 6 = 12**

### Group - E

8. (a) What is a DOI and how does it work? [[C04](Understand /IOCQ)]
- (b) Can a DOI change over time? What is the purpose of using DOIs in academic citations? How do I cite a document using its DOI? [[C04](Understand /IOCQ)]
- (c) Are DOIs only used for academic papers? [[C04](Understand /IOCQ)]
- 4 + 6 + 2 = 12**
9. (a) What is the OAIS Reference Model, and why is it important in digital preservation? [[C04](Understand /IOCQ)]

- (b) What do we mean by “persistent identification” in the context of digital objects?  
[[CO4](Understand /IOCQ)]
- (c) How does “unique identification” of digital objects differ from traditional file naming?  
[[CO4](Understand /IOCQ)]
- 4 + 6 + 2 = 12**
- 

Cognition Level	LOCQ	IOCQ	HOCQ
Percentage distribution	44.8	55.2	0