

**DATA CURATION
(CSEN 3144)**

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) A data curator typically identifies:
 - (a) Required data sets and ensures they're collected
 - (b) Data is cleansed and transformed as needed
 - (c) Making the data sets and information about them, such as their metadata and lineage documentation, available to users
 - (d) All of the above.
- (ii) Which of the following is not a part of the data curation process?
 - (a) Communication building
 - (b) Operationalize
 - (c) Model planning
 - (d) Discovery
- (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 the advantage of data sharing?
 - (a) Data sharing can violate the privacy of involved parties.
 - (b) Enables better research and development for companies and academic institutions
 - (c) Data can be misused or misinterpreted
 - (d) Enables data plagiarism and use it without copyrights.
- (v) 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
- (vi) What role does the Digital Curation Centre (DCC) play in the field of digital preservation?
 - (a) It provides information on digital curation and long-term preservation
 - (b) It is primarily a funding agency for research projects
 - (c) It focuses exclusively on the development of new storage technologies
 - (d) It serves as an international regulatory body for digital repositories.

- (vii) Which of the following best describes digital stewardship?
- (a) An initiative for digital content creation
 - (b) A technical approach to data encryption
 - (c) Holding digital resources in trust for future generations
 - (d) A process focused on economic profitability of digital content.
- (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) In the context of trusted digital repositories, what is the primary mission?
- (a) Facilitating immediate access to all digital resources
 - (b) Ensuring the deletion of outdated digital files
 - (c) Establishing digital content creation protocols
 - (d) Providing reliable, long-term access to managed digital resources.
- (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.

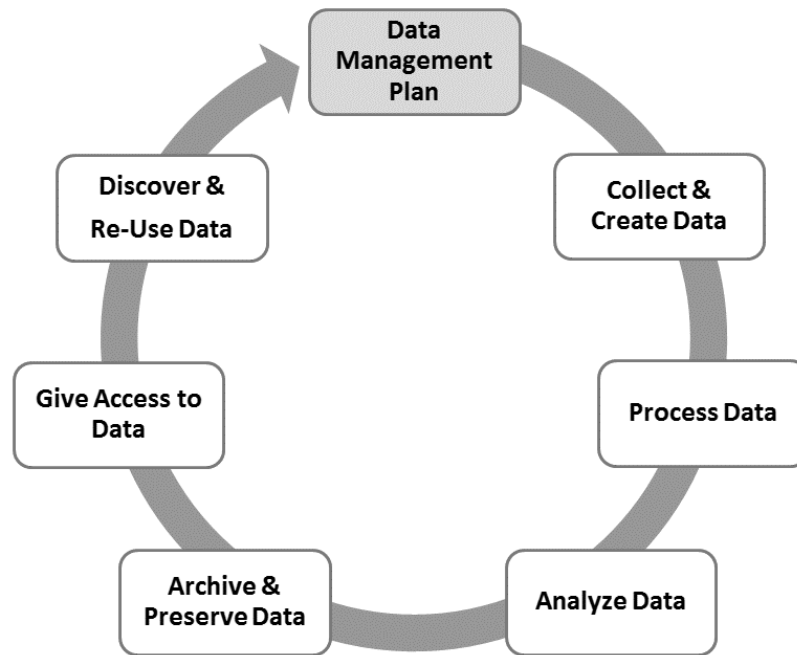
Fill in the blanks with the correct word

- (xi) _____ is the process of moving data that is no longer actively used to a separate storage device for long-term retention.
- (xii) Data _____ is a community promoting data preservation and reuse across disciplines with tools and services.
- (xiii) Digital Preservation can be defined as the _____ of digital content over time to ensure ongoing access.
- (xiv) What does 'OAIS' stands for _____.
- (xv) _____ is the process of moving data that is no longer actively used to a separate storage device for long-term retention.

Group - B

2. (a) Where does data curation fit into the data creation process? [[C01(Define/LOCQ)]]
- (b) What is the purpose of data curation? [[C01(Define/LOCQ)]]

- (c) 'Data management refers to the process of deciding and documenting how data will be collected, organized, stored and shared' - Justify this comment by the figure below: [[CO1(Define/LOCQ)]]



$$4 + 4 + 4 = 12$$

3. (a) Regarding the nature of your data, what is the subject discipline (domain, field) to which your research data relates? What is the exact nature (range, scope) of your research data? In what format(s), will you store your data in the short term after acquisition? Who owns the data arising from your research, and the intellectual property rights relating to them? [[CO4] (Understand/IOCQ)]
- (b) How your research datasets are described? [[CO2] (Understand/LOCQ)]

$$8 + 4 = 12$$

Group - C

4. (a) “a lot of big data is a lot of small data put together” – Illustrate the emerging requirements for Big Data Curation. [[CO3](Understand /IOCQ)]
- (b) Regarding the nature of your data, what is the subject discipline (domain, field) to which your research data relates? What is the exact nature (range, scope) of your research data? In what format(s), will you store your data in the short term after acquisition? Who owns the data arising from your research, and the intellectual property rights relating to them? [[CO4](Understand/LOCQ)]

$$4 + 8 = 12$$

5. (a) “The core impact of data curation is to enable more complete and high-quality data-driven models for knowledge organizations.” – Justify this statement. [[CO2](Understand/LOCQ)]
- (b) What are the important phases of data life cycle model. [[CO3] (Understand/IOCQ)]
- (c) What is the importance of data types and data formats. Give examples to support your answer. [[CO4](Understand/LOCQ)]
- (d) What is a discipline specific data format? Give an example. [[CO4](Understand/LOCQ)]

$$3 + 4 + 3 + 2 = 12$$

Group - D

6. (a) How do you choose the right metadata schema for your data needs? [[CO1](Define/LOCQ)]
 (b) What tools and standards are available for managing and validating metadata? [[CO4] (Understand/LOCQ)]
 (c) How can metadata be used to enhance data discoverability and usability? [[CO1,CO2](Understand/LOCQ)]
3 + 3 + 6 = 12
7. (a) What necessary actions are taken when a data is no longer being used? [[CO4](Understand/LOCQ)]
 (b) What is data repository? What are different types of data repository? [[CO1] (Define/LOCQ)]
 (c) Mention 3 very important criteria for archiving a data. [[CO2,CO4](Understand/IOCQ)]
4 + (2 + 3) + 3 = 12

Group - E

8. (a) How does OAIS address the challenges of digital data formats and evolving technology? [[CO4] (Understand /IOCQ)]
 (b) What are the roles of an OAIS Producer, Consumer, and Management? [[CO4] (Understand /IOCQ)]
 (c) How can organizations implement OAIS principles in a cost-effective manner? [[CO4] (Understand /IOCQ)]
4 + 6 + 2 = 12
9. (a) What is a DOI and how does it work? [[CO4] (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? [[CO4] (Understand /IOCQ)]
 (c) Are DOIs only used for academic papers? [[CO4] (Understand /IOCQ)]
4 + 6 + 2 = 12

Cognition Level	LOCQ	IOCQ	HOCQ
Percentage distribution	55.21	44.79	0