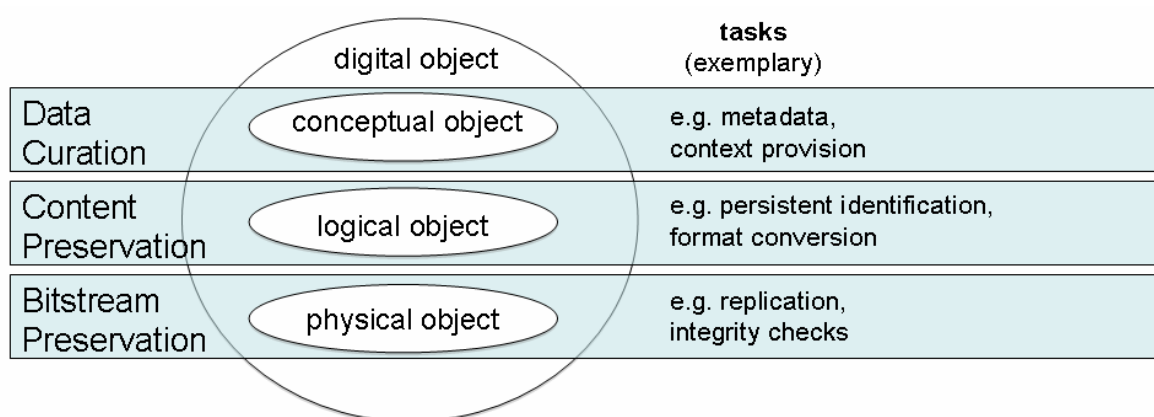


Data Curation Layers

The notions "long-term preservation" and "data curation" vary between contexts and communities. As a common terminology basis we distinguish between three curation levels that build upon each other. These three levels are directly derived from the three aspects of digital objects distinguished by the preservation community¹:

- physical - Signs Inscribed on a Medium
- logical - Processable Units
- conceptual - What We Deal with in the Real World

The following three curation levels describe - just like the analogy in digital objects - independent aspects, yet they build upon each other and may influence/support each other. Methods within a curation level can change over time as technology changes, just like an object may be reassigned to another curation level depending on its lifecycle stage. For the purposes of this terminological reference, we avoid going into technical methods or organisational responsibilities, though each of these layers clearly requires dedicated measures.



Bitstream Preservation

Bitstream Preservation ensures the integrity of each bit of a data item over time. Digital carriers are particularly in peril of media decay, and even if a carrier has a lifetime of several years or even decades, the availability of hardware readers for the carrier is often not guaranteed as we move from floppies, to CDs, to tapes, to future carriers.

Components of Bitstream Preservation include durability of the hardware carriers and readers; the number of redundant copies of data items; their distribution and independence (with regard to geography, organisation, financing, technology, etc.); as well as recurrent integrity checks.

Bitstream Preservation is primarily the responsibility of data centers and digital infrastructure.

¹ Kenneth Thibodeau: Overview of Technological Approaches to Digital Preservation and Challenges in Coming Years. CLIR Report, 2002, <http://www.clir.org/pubs/reports/pub107/thibodeau.html>

Content Preservation

Citability and accessibility of data items goes beyond the mere availability of the item's bits and touches e.g. upon the software with which it was created and its format.

Components of Content Preservation include persistent identification, technology watch (monitoring the evolution of the software environment), as well as preservation services such as format conversion, format validation, or emulation.

Institutions currently taking responsibility for Content Preservation include cultural heritage institutions such as libraries, archives, and museums, as well as some universities (or their university libraries).

Data Curation

Data Curation views the whole life cycle of digital objects and views the object in its intellectual context. It strives to preserve not only the data item, but rather the meaning of the digital object to foster comprehension and re-usability in the future when the original context (e.g. the creators, organisational environment) have disappeared.

Components of Data Curation include data and metadata modeling, appraisal of the object's value, integration of the data item in usage environments, object versioning, access control, and others. Examples include the capture, cleaning, filtering, and iterative processing and visualisation of instrumental data; or the enrichment of ancient texts with dictionaries and other contextual data from their time of creation.

Data Curation tasks usually involve considerable domain knowledge and depends on the cooperation of the creators, as well as close embedding in the creation context. Generic service providers are therefore usually not in place to fully cover Data Curation. However, more specialised service providers such as World Data Centers can adequately support curation within a community.