

## Seminario: Curación de colecciones de datos de investigación

---

Año lectivo:	<b>2017</b>
Profesor a cargo:	María Esteva
Carga horaria:	30 hs.
Modalidad:	A distancia

---

### 1. Fundamentación y objetivos

Los paradigmas de la investigación están en constante evolución. En la actualidad, la ciencia intensiva de datos y las metodologías asociadas requieren de colecciones de datos variadas, completas, bien documentadas y de gran tamaño para facilitar el análisis y el descubrimiento de patrones y fenómenos en todas las áreas científicas, de ciencias sociales y de humanidades. Tradicionalmente las investigaciones incluyen la recolección y generación de datos, pero estas colecciones, que se forman durante trabajos de campo, observaciones, encuestas, experimentos, simulaciones, etc. eran utilizadas solamente por el equipo de investigadores que volcaban los resultados de la investigación en la correspondiente publicación. Los datos obtenidos resultaban inaccesibles a otros usuarios con intenciones de re-utilizarlos en otras investigaciones. Hoy, estas colecciones son consideradas como publicaciones que dan validez y credibilidad a los trabajos de investigación y que deben ser citadas por aquellos que las reutilicen. La curación digital es el conjunto de estrategias que facilitan la creación y el mantenimiento de colecciones digitales sustentables y teniendo en cuenta su valor como material de estudio para el avance del conocimiento. En este nuevo paradigma, los profesionales de la información, en coordinación con los equipos de investigación, guían los procesos de organización, descripción y preservación de datos desde su creación durante el trabajo de investigación y hasta su publicación en repositorios digitales abiertos.

Los objetivos del seminario son:

- Conocer las teorías de curación digital.
- Conocer las prácticas y las tecnologías de punta para la curación de datos.
- Identificar los aspectos legales de la custodia y la diseminación de datos.
- Desarrollar plataformas de metadatos descriptivos, legales y de preservación de colecciones de datos.

- Diseñar flujos de curación de datos en relación a procesos activos de investigación.
- Utilizar modelos económicos para determinar el costo beneficio de la curación digital.
- Demostrar capacidad crítica, habilidad para la investigación, y capacidad para redactar informes, presentaciones, planes y trabajos de curación de datos.

## 2. Contenidos

### **Módulo 1: Los nuevos paradigmas de investigación y procesos de formación de colecciones de datos.**

1. Teorías de gestión de datos digitales
  - (a) Curación digital
  - (b) Post-custodialismo
  - (c) Ciclo de vida de los datos digitales
2. Tipos de datos digitales
  - (a) Formatos digitales y sustentabilidad
  - (b) Estructura de los datos digitales
3. Formación de colecciones de datos
  - (a) Métodos de investigación
  - (b) Flujos de trabajo
  - (c) Proveniencia de datos
  - (d) Organización de los datos

### **Módulo 2: Ciberinfraestructura para el desarrollo, curación y reutilización de datos**

1. Infraestructura técnica
  - (a) Recursos de almacenamiento, redes y sistemas de acceso
  - (b) Seguridad
  - (c) Autenticidad e integridad de los datos
2. Diseño de arquitecturas de colecciones de datos

### **Módulo 3: Metadatos para la gestión, descripción y acceso a las colecciones de datos**

1. Estándares de metadatos para datos de investigación
2. Mapeo de metadatos en arquitecturas de colección de datos

### **Módulo 4: Publicación de datos**

1. Identificadores persistentes de objetos digitales
2. Derechos de autor y licencias para distribución y uso de datos
3. Datos sensibles, confidencialidad y privacidad
4. Determinación de periodos de embargo
5. Marketing de datos

### **Módulo 5: Implementación y gestión de programas de curación en distintos escenarios**

1. Bibliotecas académicas
2. Archivos y Museos
3. Organizaciones gubernamentales
4. Ámbito científico
5. Modelos económicos para sustentar la curación de datos

### **3. Metodología de trabajo y sistema de evaluación**

Para cumplir el requisito práctico del curso, los estudiantes identificarán una colección de datos digitales de investigación para trabajarla durante el seminario. Esta puede pertenecer a un centro de investigación o ser la colección de trabajo de un investigador o equipo de investigación. La colección de datos también puede pertenecer a un archivo o museo. Como trabajos a evaluar los alumnos elaborarán:

- Proyecto de curación para la colección seleccionada: El proyecto de curación debe contemplar todos los elementos prácticos aprendidos durante el curso. Una vez

corregidos, los trabajos serán entregados a la institución/investigador cuya colección sirvió como material de investigación y práctica.

- Manual de procedimientos de curación digital: Cada alumno en el seminario tomara una de las unidades temáticas, realizara una revisión de la literatura y elaborara un comentario /análisis crítico sobre el tema. El comentario deberá reflejar realidades locales y ejemplos observados durante los trabajos prácticos. Una vez compilado y editado, el manual servirá para introducir la curación digital a científicos que desconocen la importancia del tema y para educarlos en relación a de la universidad e instituciones. De modo que todos los alumnos se interioricen de todos los aspectos del manual, la corrección, edición, y formateo, ilustración, y demás partes de la compaginación de un trabajo final se distribuirá entre los distintos alumnos.

**Notas:** El profesor evaluará los dos trabajos con notas de 1 a 10. La nota final se basará en los siguientes porcentajes:

- Trabajo de planificación de curación de una colección: 60% de la nota final (10% borrador, 50% el trabajo final)
- Sección del manual de curación a determinar hacia el final de las clases teóricas: 40% de la nota final.

En relación a la elaboración de trabajos finales, el alumno debe ser respetuoso de la propiedad intelectual de las fuentes y de las normas de comportamiento académico.

#### **4. Bibliografía**

- Arms, C. R.; Fleischhauer, C. and Murray, K., compiladores. (2013). Sustainability of Digital Formats, Planning for Library of Congress Collections.  
<http://www.digitalpreservation.gov/formats/>
- Beagrie, Neil. (2006). Digital Curation for Science, Digital Libraries, and Individuals. International Journal of Digital Curation 1, no. 1, 3-16.  
<http://dx.doi.org/10.2218/ijdc.v1i1.2>
- Beaudoin, J. E. (2011). Specters in the Archive: Faculty Digital Image Collections and the Problems of Invisibility, The Journal of Academic Librarianship, vol. 37, p. 488-494
- Becker, Christoph ... et al. (2009) . Systematic Planning for Digital Preservation: Evaluating Potential Strategies and Building Preservation Plans. International Journal on Digital Libraries 10 no. 4, 133-157.

- California Digital Library. (2012). Total Cost of Preservation: Cost Modeling for Sustainable Services. Revision 0.16. Oakland: UC Curation Center/California Digital Library, April 9.
- Caplan, Priscilla. (2009). Understanding PREMIS. Washington: Library of Congress. <http://www.loc.gov/standards/premis/understanding-premis.pdf>
- CPSR <http://www.icpsr.umich.edu/icpsrweb/landing.jsp>
- Creative Commons <http://creativecommons.org/>
- DarwinCore <http://rs.tdwg.org/dwc/>
- Data Documentation Initiative <http://www.ddialliance.org/>
- Data Management and Publishing. MIT Libraries. <https://libraries.mit.edu/data-management/>
- Data Protection and Confidentiality. The University of Edinburgh. <http://www.ed.ac.uk/records-management/data-protection>
- DataCite <https://www.datacite.org/>
- Esteva , Maria; Jordan, Christopher; Urban, Tomislav and Walling, David (2011). Cyberinfrastructure Supporting Evolving Data Collections. In Proceedings of the 8th International Conference on Preservation of Digital Objects, iPRES 2011, Singapore, November 1-4 2011. <http://ipres2011.sg/conference-proceedings>
- Esteva, Maria (2008). Formation Process and Preservation of a Natural Electronic Archive. Proceedings of the American Society for Information Science and Technology 2008 Conference, Ohio, October 24 – 29, Vol 45:1, pp 1-9. Wiley Online Library. <http://onlinelibrary.wiley.com/doi/10.1002/meet.2008.1450450270/full>
- Fishes of Texas <http://www.fishesoftexas.org/home/>
- Granda P., Thomas D, Grasso, C., Teixeira, C. (2009) Common Workflows: Health and Social Science Curation Collaborations. DigCCurr2009 Proceedings. <http://www.lulu.com/spotlight/DigCCurr2009>
- Higgins, Sarah (2008). The DCC Curation Lifecycle Model. International Journal of Digital Curation 3, no. 1, 134-140. <http://dx.doi.org/10.2218/ijdc.v3i1.48>
- iRODS: DataGrids, Digital Libraries, Persistent Archives and Real-time Data Systems. Retrieved February 3, 2013 from: <https://irods.org/>
- Ockerbloom, John Mark (2008). What Repositories Do: The OAIS Model. Everybody's Libraries, October 13, . <http://everybodyslibraries.com/2008/10/13/what-repositories-do-the-oais-model/>

- Open Planets Foundation. <http://www.openplanetsfoundation.org/>
- PaleoCore <http://paleocore.org/>
- Prodanovic, Masa; Esteva, Maria; Hanlon, Matthew; Nanda, Gaurav and Agarwal, Prateek (2015) Digital Rocks Portal: a repository for porous media images <http://dx.doi.org/10.17612/P7CC7K>
- Schreibman, Susan; Siemens, Ray; Unsworth, John eds. (2004). A Companion to Digital Humanities. Oxford: Blackwell. <http://www.digitalhumanities.org/companion/>
- Thibodeau, K. (2013) Wrestling with shape-shifters: Perspectives on preserving memory in the digital age. "The Memory of the World in the Digital Age: Digitization and Preservation, UNESCO's Memory of the World Conference. Vancouver, September 26, 2012. [http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CI/CI/pdf/mow/VC\\_Thibodeau\\_26\\_plenary1.pdf](http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CI/CI/pdf/mow/VC_Thibodeau_26_plenary1.pdf)
- Trelogan, Jessica; Rabinowitz, Adam; Esteva, Maria and Pipkin, Stephen (2010). What Do we Do with the Mess? Managing and Preserving Process History in Evolving Digital Archaeological Archives. In Contreras, F., and Melero, F.J. (eds.), Proceedings of the 38th Conference on Computer Applications and Quantitative Methods in Archaeology, Granada, Spain, April 6 – 9
- Walling, David; Esteva, Maria (2011). Automating the Extraction of Metadata From Archaeological Data Using iRods Rules. International Journal of Digital Curation, Vol 6. No 2. <http://www.ijdc.net/index.php/ijdc/article/view/196>
- 4Cproject, Collaboration to Clarify the Cost of Curation. <http://4cproject.eu/>