Curation Micro-Services: A Pipeline Metaphor for Repositories

Abrams, Stephen; Cruse, Patricia; Kunze, John; Minor, David

Curation Micro-Services: A Pipeline Metaphor for Repositories

Abstract

The effective long-term curation of digital content requires expert analysis, policy setting, and decision making, and a robust technical infrastructure that can effect and enforce curation policies and implement appropriate curation activities. Since the number, size, and diversity of content under curation management will undoubtedly continue to grow over time, and the state of curation understanding and best practices relative to that content will undergo a similar constant evolution, one of the overarching design goals of a sustainable curation infrastructure is flexibility. In order to provide the necessary flexibility of deployment and configuration in the face of potentially disruptive changes in technology, institutional mission, and user expectation, a useful design metaphor is provided by the Unix pipeline, in which complex behavior is an emergent property of the coordinated action of a number of simple independent components. The decomposition of repository function into a highly granular and orthogonal set of independent but interoperable micro-services is consistent with the principles of prudent engineering practice. Since each micro-service is small and self-contained, they are individually more robust and collectively easier to implement and maintain. By being freely interoperable in various strategic combinations, any number of micro-services-based repositories can be easily constructed to meet specific administrative or technical needs. Importantly, since these repositories are purposefully built from policy neutral and protocol and platform independent components to provide the function minimally necessary for a specific context, they are not constrained to conform to an infrastructural monoculture of prepackaged repository solutions. The University of California Curation Center has developed an open source micro-services infrastructure that is being used to manage the diverse digital collections of the ten campus University system and a number of non-university content partners. This paper provides a review of the conceptual design and technical implementation of this micro-services environment, a case study of initial deployment, and a look at ongoing micro-services developments.