

SIDR-UEVE: a pilot for eliciting a French academic policy on data preservation in Life sciences

Emmanuel Dequier¹, Morad Mechref², Alain Zasadzinski², Marie-Christine Jacquemot², Florian Mazur², Herbert Gruttemeier², Philippe Savarin³, Marie-Jeanne Clément³, Olek Maciejak³, Delphine Pflieger⁴, Florence Gonnet⁴, Regis Daniel⁴, William Buchmann⁴, Patrick Curmi³, Jeanine Tortajada⁴, & Magali Roux²

¹Genopole, Global Infrastructures and Platforms, 5 rue Henri Desbruères 91030 Evry Cedex, France, e-mail: Emmanuel.dequier@genopole.fr

²INIST, 2 Allée du parc-de-Brabois, 54519 Vandoeuvre-lès-Nancy, France, e-mail: {surname.name}@inist.fr

³SABNP UMR S 829 Inserm - University of Evry Val d'Essone (UEVE), Bd. F. Mitterrand, 91025 Evry Cedex, France, e-mail: {surname.name}@univ-evry.fr

⁴LAMBE, UMR 8587, CNRS - CEA - University of Evry Val d'Essone (UEVE), Bd. F. Mitterrand, 91025 Evry Cedex, France, e-mail: {surname.name}@univ-evry.fr

Natural sciences dealing with data-intensive production and sharing of distributed online resources are known as e-sciences. Although based on classical observational and/or experimental approaches, they use complex equipments that require extra-organizational research schemes and dedicated infrastructures. For example, in Biology and Chemistry, high-throughput facilities produce large volumes of data that need special methods and tools to be tackled safely and consistently. To achieve this, each facility must have its own in-house manpower to insure a 24-hour watchfulness maintenance including the management of disaster operations, if any. Otherwise, the data-producing platforms might delegate data administration to external committed infrastructure that will guarantee data security as well in terms of confidentiality and intellectual property as in physical preservation.

To achieve this, we are developing a pilot named SIDR-UEVE to examine main aspects and conditions for the preservation and the accessibility of data produced from academic UEVE facilities. SIDR is a national data repository under development to insure the collect, the preservation and the diffusion of Omics research data in Life sciences that include a scientific and technical Board-approved access plan. In SIDR, metadata for transcriptomics, proteomics, etc., are collected according to the standards developed by the corresponding communities and articulated to the Investigation-Study-Assay (ISA) core designed at the European Bioinformatics Institute (EBI). This standards-based approach makes the SIDR repository highly interoperable with other data infrastructures all over the world.

Accordingly, the SIDR-UEVE pilot is the SIDR extension to Mass Spectrometry data and will be broadened to Nuclear Magnetic Resonance ones. The pilot focuses not only on important issues concerning technical aspects (for example, the authentication procedures and tools) but also on legal principles according to platform stakeholders (for example, intellectual properties, data restitution/diffusion/deletion, etc). SIDR-UEVE relationships scenarios are designed that should feed the eliciting of a French data preservation policy in Life sciences.
