การเก็บถาวรอิเล็กทรอนิกส์ : ข้อสังเกตซอฟต์แวร์ในการจัดการคลังเก็บเอกสารอิเล็กทรอนิกส์

สุภาพร ชัยธัมมะปกรณ์

บทความนี้เป็นการนำเสนอเครื่องมือในการจัดการคลังเก็บเอกสารอิเล็กทรอนิกส์ที่เป็นที่รู้จักกันอย่างแพร่หลาย คือ ล็อคคิส อีพรินท์ และดีสเปซ โดยจะพูดถึงในแง่ของผู้ใช้ ค่าใช้จ่าย เทคโนโลยีขั้นพื้นฐาน เทคนิคพิเศษที่จำเป็นต้องใช้ และการใช้งานต่างๆ ห้องสมุดใดที่จะนำเครื่องมือไปใช้ในการจัดเก็บถาวรสารสนเทศอิเล็กทรอนิกส์ต้องใช้เวลา การวางแผน และความรู้ทางเทคนิค ซึ่งขึ้นอยู่กับว่าห้องสมุดสามารถนำเครื่องมือที่เหมาะสมมาใช้ได้ตรงกับความต้องการและตรงกับทรัพยากรหรือไม่

E-Archiving : An Overview of Some Repository Management Software Tools

Supaporn Chaithammapakorn

The article presents LOCKSS, EPrints and DSpace which are the most widely know repository management tools in terms of who uses them, their cost, underlying technology, the required know how, and functionalities. Any archiving project requires time, planning, and technical know-how it is up to libraries to match the right tool to its needs and resources.

การเก็บถาวรอิเล็กทรอนิกส์ : ข้อสังเกตซอฟต์แวร์ในการจัดการคลังเก็บเอกสารอิเล็กทรอนิกส์ *

สุภาพร ชัยธัมมะปกรณ์, ผู้แปล **

บทนำ

ในระยะหลังมีการริเริ่มพัฒนาซอฟต์แวร์สำหรับการจัดการคลังเก็บเอกสารอิเล็กทรอนิกส์ขึ้นเป็นจำนวนมาก บางองค์กรหรือหน่วยงานจัดทำขึ้นเพื่อใช้ในการสงวนรักษาเนื้อหาที่อาจจะสูญหายได้ บางหน่วยงานจัดทำขึ้นเพื่อให้สามารถเข้าถึงทรัพยากรสารสนเทศได้มากขึ้น ความเคลื่อนไหวที่ให้ใช้ได้อย่างเปิดกว้างนี้เป็นปัจจัยสำคัญที่ทำให้เกิดการพัฒนานี้ขึ้น การริเริ่มเอกสารดิจิทัล เช่น เอกสารก่อนการจัดพิมพ์ (pre-print) เอกสารหลังการจัดพิมพ์ (post-print) และเซิร์ฟเวอร์สำหรับจัดเก็บเอกสาร (document servers) ถูกสร้างขึ้นมาเพื่อเป็นแนวทางใหม่ในวงการพิมพ์ เนื่องจากยังไม่สามารถควบคุมราคาของวารสารได้ โดยเฉพาะวารสารของหน่วยงานทางด้านวิทยาศาสตร์ เทคโนโลยี และทางการแพทย์ อีกทั้งจำนวนผู้เขียนและมหาวิทยาลัยที่มีความกระตือรือร้นที่จะมีส่วนในการพิมพ์และสงวนรักษาด้วยตนเองมีมากขึ้นๆ

ในการเลือกเครื่องมือ ห้องสมุดต้องพิจารณาถึงประเด็นต่างๆ ดังนี้

- ประเภทของทรัพยากรสารสนเทศที่ต้องการเก็บ
- ระยะเวลาของการเก็บ
- ความเหมาะสมของซอฟต์แวร์
- ค่าใช้จ่ายในการติดตั้งระบบ
- ความรู้ด้านเทคนิคพิเศษ

ในบทความนี้จะพูดถึงโปรแกรมที่ช่วยสนับสนุนงานด้านนี้สามตัวด้วยกันคือ ล็อคคิส (LOCKSS) อีพริ้นท์ (EPrints) และ ดีสเปซ (DSpace) ซึ่งเป็นเครื่องมือในการจัดการการเก็บที่เป็นที่รู้จักกันอย่างแพร่หลาย โดยจะพูดถึงทั้งในแง่ของผู้ใช้ ค่าใช้จ่าย เทคโนโลยีพื้นฐาน เทคนิคพิเศษที่จำเป็นต้องใช้ และหน้าที่การทำงานต่างๆ

---

* เรียบเรียงจาก Prudlo, Marion. "E-Archiving: an overview of some repository management software tools." ARIADNE, issue 43 (April 2005). Available: http://www.ariadne.ac.uk/issue43/prudlo/ 7/15/2005 6:30.

** บรรณารักษ์ชำนาญการระดับ 8 ฝ่ายวิเคราะห์ทรัพยากรสารสนเทศ หอสมุดกลาง สถาบันวิทยบริการ จุฬาลงกรณ์มหาวิทยาลัย

ล็อคคิส (LOCKSS)

โดยทั่วไปห้องสมุดจะไม่ได้ซื้อตัวเนื้อหาของวารสารอิเล็กทรอนิกส์แต่จะได้เป็นสิทธิการเข้าใช้ (license) ได้ในช่วงเวลาหนึ่ง ซึ่งหากไม่ต่ออายุสมาชิกจะไม่สามารถใช้เนื้อหานั้นได้ ดังนั้นหากทุกห้องสมุดขาดอายุสมาชิกไป หรือผู้จัดทำ/ผู้พิมพ์เลิกกิจการไป เนื้อหาที่ว่าจะไม่มีให้ใช้ได้อีกต่อไป ก่อนที่จะมาถึงยุควารสารอิเล็กทรอนิกส์ ห้องสมุดต้องเป็นสมาชิกวารสารฉบับตีพิมพ์ด้วยตนเองเท่านั้น เนื่องจากไม่มีวิธีอื่นที่สะดวกและเร็วกว่านี้ในการเข้าถึงวารสาร ปัจจุบันห้องสมุดไม่จำเป็นต้องบอกรับวารสารฉบับตีพิมพ์ทุกฉบับอีกต่อไปแล้ว เพราะห้องสมุดสามารถเข้าถึงวารสารได้จากฐานข้อมูลและการบอกรับเป็นสมาชิกวารสารอิเล็กทรอนิกส์ การเป็นสมาชิกวารสารฉบับตีพิมพ์ หมายถึง ห้องสมุดได้เป็นเจ้าของวารสารฉบับนั้นๆ อย่างถาวรตลอดไป ดังนั้น ผลของการที่วารสารฉบับตีพิมพ์ที่มีห้องสมุดแห่งใดแห่งหนึ่งเป็นเจ้าของก็คือ จะมีห้องสมุดใดห้องสมุดหนึ่งในประเทศสหรัฐอเมริกา หรือในที่อื่นๆ เก็บรักษาวารสารเล่มนั้นไว้ด้วยการเย็บเล่มหรือถ่ายเป็นไมโครฟิล์มและให้ยืมผ่านระบบการยืมระหว่างห้องสมุด จึงเป็นการเก็บรักษาเอกสารอย่างถาวรไปโดยปริยาย ซึ่งระบบสมาชิกแบบอิเล็กทรอนิกส์ไม่อาจเก็บรักษาในลักษณะเดียวกันได้

ล็อคคิส (LOCKSS ย่อมาจาก Lots of Copies Keep Stuff Safe) ได้รับการพัฒนาขึ้นที่มหาวิทยาลัยสแตนฟอร์ด (Standford University) โปรแกรมนี้ จะช่วยให้ห้องสมุดสามารถให้บริการเนื้อหาของวารสารอิเล็กทรอนิกส์ นับตั้งแต่วันที่ห้องสมุดเป็นสมาชิก โดยห้องสมุดจะจัดเก็บถาวรไว้ ซึ่งทำให้ยังคงให้บริการเนื้อหาต่อไปได้ แม้ว่าหมดอายุสมาชิกแล้วก็ตาม เนื่องจากล็อคคิสเป็นระบบที่รวบรวมเนื้อหาใหม่อย่างช้าๆ จึงเหมาะสำหรับการเก็บถาวรของเนื้อหาที่คงที่ (static) ไม่เปลี่ยนแปลงบ่อย เพราะฉะนั้นจุดประสงค์เบื้องต้นของล็อคคิส คือ การสงวนรักษาการเข้าถึงวารสารอิเล็กทรอนิกส์ เนื่องจาก เนื้อหาของวารสารจะถูกเพิ่มเติมเข้ามาตามคาบที่แน่นอน (regular intervals) ตัวต้นฉบับจะถูกเก็บรักษาแยกไว้ แทนการสร้างฉบับสำรอง (back-up copy) เพื่อให้มั่นใจในความเชื่อถือได้ของเนื้อหา และเพื่อเป็นหลักประกันว่าจะคงมีอยู่ตลอดไปโดยมีการเก็บสำเนาไว้ถึงประมาณ 6 ชุด

ล็อคคิสเป็นระบบเปิดสำหรับห้องสมุดใดก็สามารถใช้ได้ ปัจจุบันมีสถาบันเกือบ 100 แห่งทั่วโลกที่ใช้โปรแกรมล็อคคิส ซึ่งส่วนใหญ่เป็นหน่วยงานในประเทศสหรัฐอเมริกาและประเทศต่างๆ ในยุโรป รวมทั้งระบบสนับสนุนการพิมพ์ (publishing platform) ต่างๆ ด้วย เช่น โปรเจ็คมิวส์ (Project Muse) สำนักพิมพ์แบล็คเวลล์ (Blackwell Publishers) เอ็มเมอรัลกรุ๊ปพับบลิชชิง (Emerald Group Publishing) เนเจอร์พับบลิชชิงกรุ๊ป (Nature Publishing Group) และครูเวอร์อะคาเดมิคพับบลิชเชอร์ (Kluwer Academic Publishers) นอกจากนี้ผู้ใช้ยังมีวารสารอีกจำนวนหนึ่งที่ถูกจัดเก็บเป็นการถาวรที่ผู้สนใจจะเรียกอ่านได้จากเว็บไซต์โดยไม่ต้องเสียค่าใช้จ่ายใดๆ ทั้งสิ้น

ล็อคคิสจะจัดเก็บสิ่งพิมพ์ที่มีกำหนดการพิมพ์ที่แน่นอน และส่งออกไปผ่านเอชทีทีพีที่มียูอาร์แอล แต่สิ่งพิมพ์ เช่น เว็บไซต์นั้นมีการเปลี่ยนแปลงอยู่ตลอดเวลาจึงไม่เหมาะสำหรับการเก็บถาวรด้วยล็อคคิส ถ้าวารสารฉบับใดมีโฆษณาที่เปลี่ยนไป โฆษณานั้นจะไม่ถูกเก็บรักษาไว้ เมื่อไม่นานมานี้มีการสำรวจว่า สามารถใช้ล็อคคิสมาใช้จัดเก็บสิ่งพิมพ์รัฐบาล ที่ถูกตีพิมพ์บนเว็บได้หรือไม่ และมีการริเริ่มอีกโครงการหนึ่งคือ การนำล็อคคิสมาใช้เป็นหน่วยเก็บถาวรของเว็บไซต์ที่ไม่มีการเปลี่ยนแปลงอีกต่อไป

ข้อดีของการสงวนรักษาเนื้อหาด้วยล็อคคิส คือ ราคาถูกและไม่ต้องใช้เวลามากนัก ห้องสมุดที่ต้องการมีส่วนร่วมในการใช้ล็อคคิส ต้องการเพียงเครื่องคอมพิวเตอร์ที่ทำหน้าที่เป็นเครื่องเสมือน ซึ่งใช้คอมพิวเตอร์ทั่วไปที่ราคาไม่แพง คอมพิวเตอร์นี้จะต้องเชื่อมต่อกับอินเทอร์เน็ตได้ แม้ว่าการติดต่ออินเทอร์เน็ตด้วยการหมุนโทรศัพท์จะไม่ดีพอก็ตาม ความต้องการของระบบต่ำสุดก็คือ ซีพียูอย่างน้อย 600 เมกกะเฮิร์ตซ์ แรมอย่างน้อย 128 เมกกะไบต์ และมีดิสก์ไดรฟ์ 1 หรือ 2 ตัว ที่มีความจุอย่างน้อย 60 กิกะไบต์ โดยโปรแกรมล็อคคิสสามารถเรียกได้จากซีดีซึ่งประกอบด้วยระบบปฏิบัติการโอเพ่นบีเอสดี (OpenBSD) รายละเอียดในการลงโปรแกรมทีละขั้นตอนสามารถดาวน์โหลดได้จากเว็บไซต์ของล็อคคิสที่ http://www.lockss.org/publidocs/install.html ผู้ที่ลงโปรแกรมควรมีทักษะและประสบการณ์ในการตั้งค่าต่างๆ ของซอฟต์แวร์ เมื่อติดตั้งล็อคคิสเสร็จเรียบร้อยแล้วมันจะทำงานโดยอัตโนมัติ ผู้ดูแลระบบคอยเฝ้าสังเกตเพียงเล็กน้อยเท่านั้น สำหรับการสนับสนุนในด้านเทคนิค หน่วยงานหรือองค์กรสามารถเข้าร่วมกับกลุ่มผู้ใช้โปรแกรมล็อคคิส (LOCKSS Alliance) ได้

ล็อคคิสเก็บเนื้อหาของวารสาร โดยการเข้าไปเก็บจากเว็บไซต์ ของสำนักพิมพ์ต่างๆ อย่างค่อยเป็นค่อยไปและสม่ำเสมอ รูปแบบที่ล็อคคิสสามารถเก็บได้ คือ เอชทีเอ็มแอล (HTML) เจเพ็ค (jpg) จิฟ (gif) เวฟ (waf) พีดีเอฟ (pdf) ตัวโปรแกรมใส่เมทาดาตาจากสำนักพิมพ์มากกว่าเมทาดาตาที่ใส่จากห้องสมุด ห้องสมุดมีทางเลือกในการให้สร้างเมทาดาตาอยู่ในส่วนของผู้ดูแลระบบ เมื่อถูกเรียกใช้งานแคช (cache) จะทำหน้าที่แจกจ่ายเนื้อหาเสมือนเป็นเว็บพร็อกซี (web proxy) ระบบจะไปค้นหาวารสารจากเว็บไซต์ของสำนักพิมพ์ก่อน หรือถ้าหาไม่ได้ก็จะไปหาจากแคชให้ การเข้าไปเก็บจากเว็บไซต์ของสำนักพิมพ์เพื่อ ต้องการให้สถาบันต่างๆ ผ่านการอนุญาตจากสำนักพิมพ์ก่อน ซึ่งการได้รับอนุญาตนี้เป็นการได้รับผ่านข้อตกลงการอนุญาตให้เข้าใช้ (licence agreement) แล้วสำนักพิมพ์จะแจ้งในเว็บไซต์ของสำนักพิมพ์ ว่าในวารสารแต่ละหัวมีฉบับใดบ้างและฉบับใดได้รับอนุญาตให้ล็อคคิสเก็บข้อมูลไปได้

เนื่องจากวารสารแต่ละหัวมีคุณสมบัติเฉพาะ จึงต้องมีซอฟต์แวร์เสริม (plug-in) มาช่วย เช่น ช่วยในการหาวารสาร หรือตรวจสอบกำหนดการออกของวารสาร เป็นต้น

ลักษณะที่จำเป็นของการเก็บถาวรอิเล็กทรอนิกส์คือ การทำให้แน่ใจว่าทรัพยากรยังคงมีอยู่ เชื่อถือได้ และไม่มีข้อผิดพลาดใดๆ ล็อคคิสจึงมีกระบวนการตรวจสอบความผิดพลาดของเนื้อหา และมีการทำสำรองข้อมูลอย่างอัตโนมัติ มีการเปรียบเทียบเนื้อหาอย่างต่อเนื่องระหว่างวารสารในแคชเอง และวารสารหัวเดียวกันในแคชตัวอื่นๆ โดยนำระบบการหยั่งข้อมูล (polling system) มาใช้ ซึ่งกระบวนการนี้สำเร็จลุล่วงได้โดยการใช้แอลแคพ (LCAP หรือ Library Cache Auditing Protocol) คือ ถ้าเนื้อหาระหว่างวารสาร 2 ฉบับไม่ตรงกัน ระบบจะจัดการดาวน์โหลดฉบับใหม่ที่สมบูรณ์กว่า (fresh copy) จากสำนักพิมพ์หรือจากคอมพิวเตอร์ตัวอื่นของล็อคคิส ด้วยกระบวนการนี้เองล็อคคิสสามารถซ่อมความเสียหายบางอย่างได้ ระบบการหยั่งข้อมูลนี้ทำให้มั่นใจได้ว่าเนื้อหาได้รับการสงวนรักษาอย่างน่าเชื่อถือ มีการแก้ไขข้อผิดพลาดและมีการซ่อมข้อมูลที่หายไป ถ้าเว็บไซต์ของสำนักพิมพ์ไม่สามารถให้บริการได้ล็อคคิสก็สามารถให้บริการได้จากแคชที่เก็บไว้โดยที่ผู้ใช้ไม่ทราบเลย

นอกจากนี้ ระบบการหยั่งข้อมูลสามารถป้องกันการบุกรุกจากไวรัสหรือความเสียหายจากสาเหตุอื่นๆ ได้อีกด้วย เนื่องจากระบบการหยั่งข้อมูลนี้สามารถจับการโจมตีได้แต่มนุษย์ต้องเข้าไปป้องกันความเสียหายนั้น และถึงแม้ระบบจะถูกทำลาย ล็อคคิสก็ยังมีสถาปัตยกรรมแบบกระจาย (decentralized architecture) ซึ่งเป็นมาตรการรักษาความปลอดภัยในตัวเอง ตราบเท่าที่ยังไม่มีจุดใดจุดหนึ่งเสียเนื่องจากทางกายภาพของแคช

ล็อคคิสไม่ห่วงในเรื่องเครื่องมือที่ใช้เพื่อการเก็บถาวร เพราะถ้าฮาร์ดแวร์ล้าสมัย เนื้อหาที่แคชไว้แล้ว สามารถย้ายไปยังตัวใหม่ เทคโนโลยีที่เปลี่ยนแปลงไป เช่น รูปแบบของแฟ้มข้อมูลเปลี่ยน อาจทำให้ไม่สามารถอ่านสารสนเทศอิเล็กทรอนิกส์ได้ ล็อคคิสได้เริ่มพัฒนาระบบที่ทำให้ส่งข้อมูลที่เก็บไว้จากรูปแบบหนึ่งไปสู่อีกรูปแบบหนึ่งได้ โดยการผ่านวิธี "migration on access" ซึ่งหมายถึงเนื้อหาจะถูกรักษาไว้ในรูปแบบของมันเอง จนกระทั่งเมื่อมีผู้อ่านเรียกใช้จึงถูกแปลงไปสู่ระบบใหม่ที่ใช้ในปัจจุบัน ในเดือนมกราคม ปี ค.ศ. 2005 คณะพัฒนาของล็อคคิสได้ตีพิมพ์บทความแสดงรายละเอียดของต้นแบบ ซึ่งคณะพัฒนาประสบผลสำเร็จในการย้ายจากรูปแบบจิฟ (gif) เป็นพีเอ็นจี (PNG) ได้ และเมื่อไม่นานมานี้ล็อคคิสกำลังพัฒนาตัวแปลง (converter) รูปแบบ ซึ่งจะอำนวยความสะดวกในการย้ายรูปแบบในวงที่กว้างกว่าเดิม และเพื่อพัฒนากระบวนการนี้ ล็อคคิสกำลังวางแผนเพื่อออกมาตรฐานของรูปแบบที่สมบูรณ์ และสามารถดึงส่วนที่เป็นข้อมูลทางบรรณานุกรมเมทาดาตาออกมา

อีพรินท์ (EPrints)

จำนวนผู้เขียนที่เพิ่มขึ้นอย่างต่อเนื่อง ทำให้เกิดการเลี่ยงการใช้สำนักพิมพ์จัดพิมพ์แบบเดิมๆ มาเป็นการจัดพิมพ์ด้วยตนเอง ผู้เขียนบางคนส่งงานของตนไปเก็บไว้ที่เซิร์ฟเวอร์เพื่อเก็บงานก่อนการพิมพ์ (preprint) เช่น arXiv.org (http://arxiv.org) CERN Document Server (http://cdsweb.cern.ch/?c=Preprints) หรือ Cogprints (http://cogprints.ecs.soton.ac.uk) แต่มีบ้างที่นำผลงานไว้ในเซิร์ฟเวอร์ที่เก็บงานหลังการพิมพ์ (post-print) เพื่อที่จะให้สำนักพิมพ์ได้ใช้บทความไประยะหนึ่งก่อนคืนลิขสิทธิ์มาให้ผู้เขียนหนังสือ มีสถาบันจำนวนหนึ่งจัดทำเป็นการสงวนรักษาและจัดหาช่องทางให้การเข้าถึงวรรณกรรมที่ไม่ตีพิมพ์เผยแพร่ (grey literature) ของสถาบันนั้นๆ เช่น GrayLIT Network (http://www.graylit.osti.gov/) ETH E-Collection (http://e-collection.ethbib.ethz.ch/index_e.html) หรือ SPIRES (http://www.slac.standford.edu/spires/hep/) คลังเก็บบางแห่งเป็นแบบลูกผสม (hybrid) ที่เก็บมากกว่าสารสนเทศประเภทเดียว

อีพรินท์ เป็นเครื่องมือทางซอฟต์แวร์ ที่ใช้ในการจัดการการเก็บรักษาของงานวิจัย ในรูปของหนังสือ โปสเตอร์ หรือ เอกสารการประชุม มีวัตถุประสงค์ไม่ใช่เพื่อการเก็บรักษาในระยะยาว แต่เพื่อให้มั่นใจได้ว่าสามารถอ่านและเข้าถึงสารสนเทศ ผ่านเทคโนโลยีที่เปลี่ยนแปลงไปโดยให้สถาบันเก็บ รวบรวม และจัดหาเว็บไซต์เพื่อเป็นการเข้าถึงสารสนเทศให้ ปัจจุบัน มีคลังเก็บมากกว่า 140 แห่งที่ใช้อีพรินท์ในการจัดการ ตัวอย่างเช่น มหาวิทยาลัยควีนส์แลนด์ ในประเทศออสเตรเลียใช้อีพรินท์เพื่อเป็นที่เก็บรายงานการวิจัยเสมือนเป็นตู้แสดงงานวิจัยที่เป็นผลงานของบุคลากรสายวิชาการของมหาวิทยาลัยควีนส์แลนด์และนักศึกษาระดับหลังปริญญาโท ซึ่งเป็นสิ่งพิมพ์ที่มีทั้งก่อนและหลังการประเมินโดยคนในวงการ (peer review) (http://eprint.uq.edu.au/) และที่มหาวิทยาลัยพิตต์สเบอร์กมี PhiSci Archive เพื่อเก็บรักษาผลงานก่อนการพิมพ์ ในสาขาปรัชญาของวิทยาศาสตร์ (http://philsci-archive.pitt.edu)

อีพรินท์เป็นโอเพนซอส ได้รับการพัฒนาโดยมหาวิทยาลัยเซาท์แฮมตันในประเทศสหราชอาณาจักร (http://software.eprints.org) เป็นระบบที่ใช้เป็นโครงการริเริ่มเก็บถาวรแบบเปิด (OAI - Open Archives Initiative) ซึ่งทำให้สามารถสืบค้นข้ามระหว่างหน่วยเก็บถาวรได้ เมื่อได้มีการลงทะเบียนกับโครงการริเริ่มเก็บถาวรแบบเปิด คลังเอกสารของสมาชิกนั้นจะถูกรวบรวมโดยอัตโนมัติเพื่อเก็บเมทาดาตารวมเข้าเป็นคลังเก็บทั่วโลกนั้น ทำให้เป็นการเพิ่มคุณค่าการให้บริการหน่วยงานทางวิชาการและวิทยาศาสตร์ทั่วโลก

ปัจจุบันอีพรินท์มีรุ่นล่าสุดคือ 2.3.13 (พ.ศ. 2546) การติดตั้งระบบ สามารถทำตามที่ระบบกำหนดเอาไว้แล้ว (default) ซึ่งช่วยให้เร็วและสะดวกขึ้น โดยไม่ต้องใช้ทักษะทางเทคนิคมากนัก

แต่ก็ควรมีทักษะในโปรแกรมอาปาเช (Apache) มายเอสคิวแอล (MySQL) เพิร์ล (Perl) และเอ็กซ์เอ็มแอล (XML) ผู้ดูแลระบบติดตั้งซอฟต์แวร์ลงบนเซิร์ฟเวอร์ สั่งให้ทำงานด้วยสคริปท์และมีการดูแลบ้างเท่านั้น

อีพรินท์สามารถทำงานบนลีนุกซ์ (Linux) โซลาริส (Solaris) หรือบนแมคโอเอสเอ็กซ์ (MacOS/X) เว็บเซิร์ฟเวอร์ที่ลงโปรแกรมอาปาเช (Apache Web server) โดยมีฐานข้อมูลที่ใช้โปรแกรมมายเอสคิวแอล ในด้านการสนับสนุนทางด้านเทคนิคนั้น อีพรินท์มีเมล์ลิ่งลิสที่สามารถสมัครเป็นสมาชิกได้

อีพรินท์มีส่วนต่อประสานกับผู้ใช้ (user interface) ที่สามารถจะปรับแต่งตามความต้องการได้ ผู้เขียนที่ต้องการส่งผลงานต้องลงทะเบียนก่อนโดยต้องบอกถึงประเภทของงานที่ส่ง (หนังสือ วิทยานิพนธ์ ฯลฯ) รวมทั้งลงเมทาดาตาไม่ว่าจะเป็นเค้าร่าง (schema) ใดก็ได้ ซึ่งขึ้นอยู่กับผู้ดูแลตัดสินว่าประเภทของงานใดควรจะถูกจัดเก็บ รวมทั้งเมทาดาตาองค์ประกอบใดบ้างที่ควรลงตามประเภทของงาน รายการใดที่ถูกส่งเข้าไป เจ้าของผลงานจะต้องถูกตรวจสอบ และสารสนเทศที่ถูกส่งขึ้นไปนี้จะถูกเก็บไว้ใน 'ที่พัก' (buffer) ก่อนเพื่อได้รับการประเมินจากเจ้าหน้าที่ ก่อนจะจัดเก็บในระบบในขั้นสุดท้าย

ผู้ใช้สามารถใช้ทางเลือกในการเรียกดูรายการจากหัวเรื่อง ชื่อผู้เขียน ปีที่พิมพ์ ประเภท หรือการเพิ่มเติมครั้งล่าสุด นอกจากนี้ยังสามารถสืบค้นได้จากเขตข้อมูล เช่น ชื่อเรื่อง สาระสังเขป หรือเนื้อหาฉบับเต็ม เขตข้อมูลที่สามารถเป็นตัวกำหนดให้ค้นได้ขึ้นอยู่กับผู้ดูแลระบบ ตัวอย่างของส่วนต่อประสานกับผู้ใช้ของอีพรินท์ สามารถหาดูได้จาก Cogoprints archive (http://cogprints.ecs.soton.ac.uk/) การแสดงผลการสืบค้นประกอบด้วยชื่อผู้แต่ง ปีพิมพ์ ชื่อเรื่อง สำนักพิมพ์ และจำนวนหน้า ผู้ใช้สามารถเข้าถึงข้อมูลโดยการอ่านเนื้อหาฉบับเต็มหรืออ่านสาระสังเขปก่อนได้โดยมีหัวเรื่องและคำสำคัญประกอบไว้ด้วย ที่มหาวิทยาลัยควีนส์แลนด์เทคโนโลยีในประเทศออสเตรเลีย ผู้เข้าใช้และผู้มีส่วนร่วมสามารถเห็นสถิติการเข้าใช้ด้วย (http://eprints.qut.edu.au/)

ดีสเปซ (DSpace)

ดีสเปซเป็นโอเพนซอส ได้รับการพัฒนาโดยห้องสมุดสถาบันเทคโนโลยีแห่งแมสซาชูเซตส์และบริษัทฮิวเลตต์-แพคการ์ด มีรุ่นล่าสุดคือ 1.2.2 (วันที่ 4 พฤษภาคม พ.ศ. 2546)

จากเว็บไซต์ของดีสเปซ (http://sourceforge.net/projects/dspace/) ได้ระบุว่าสถาบันหรือหน่วยงานสามารถนำซอฟต์แวรนี้ไปใช้ทำงานได้ดังนี้

- เก็บและอธิบายสารสนเทศดิจิทัลโดยการใช้กระแสงาน (workflow)
- เผยแพร่สารสนเทศดิจิทัลของสถาบันผ่านทางเว็บไซต์ เพื่อให้ผู้ใช้สามารถสืบค้นและค้นคืนรายการที่ต้องการได้
- สงวนรักษาสารสนเทศดิจิทัลในระยะยาว

มีหน่วยงานมากกว่า 100 แห่งนำดีสเปซไปใช้ เช่น ห้องสมุดดิจิทัลซิสซา (Sissa Digital Library) เป็นตัวอย่างของคลังเอกสารดิจิทัล ที่เป็นภาษาอิตาเลียน (https://digitallibrary.sissa.it/index.jsp) ซึ่งประกอบด้วยผลงานก่อนตีพิมพ์ รายงานทางวิชาการ (technical report) รายงานที่กำลังดำเนินการ (working papers) และรายงานการประชุม (conference papers) มหาวิทยาลัยเกน (Universiteit Gent) ในประเทศเบลเยี่ยม ใช้ดีสเปซในการเก็บถาวร รูปภาพ เช่น รูปภาพ ภาพพิมพ์ (prints) ภาพวาด (drawings) และแผนที่ (https://archive.ugent.be/handle/1854/219) และที่สถาบันเทคโนโลยีแห่งแมสซาชูเซตส์เอง มีคลังเก็บเอกสารดิจิทัลขนาดใหญ่ สำหรับเก็บสารสนเทศประเภท ผลงานก่อนตีพิมพ์ รายงานทางวิชาการ รายงานที่กำลังดำเนินการ และภาพต่างๆ (https://dspace.mit.edu/imdex.jsp)

โปรแกรมดีสเปซเป็นโปรแกรมที่มีความยืดหยุ่น มากกว่าโปรแกรมอีพรินท์ เนื่องจากดีสเปซมีจุดประสงค์ในการเก็บถาวร เนื้อหาที่มีความหลากหลาย เช่น บทความ ชุดข้อมูล (data sets) รูปภาพ แฟ้มข้อมูลเสียง (audio files) แฟ้มข้อมูลวีดิทัศน์ (video files) แฟ้มข้อมูลคอมพิวเตอร์ (computer files) และทรัพยากรสารสนเทศดิจิทัลอื่นๆ ของห้องสมุด ดีสเปซเป็นโปรแกรมตัวแรกในการเก็บถาวรเว็บไซต์ โดยทำหน้าที่เก็บตัวของมันเอง (storing self-contained) เอกสารเอชทีเอ็มแอลที่เป็นประเภทคงที่ และดีสเปซยังเป็นสมาชิกในโครงการริเริ่มเก็บถาวรแบบเปิด (OAI - Open Archives Initiative) อีกด้วย

ดีสเปซเป็นโปรแกรมที่เหมาะสำหรับองค์กรขนาดใหญ่ที่มีความซับซ้อนมาก คือ มีการส่ง (submit) สารสนเทศจากหลายๆ หน่วยงานที่มีความแตกต่างกัน ซึ่งเรียกว่า ชุมชน (community) เนื่องจากสถาปัตยกรรมของระบบเลียนแบบมาจากโครงสร้างขององค์กรที่ใช้ดีสเปซนั่นเอง

ดีสเปซทำงานบนยูนิกซ์ หรือระบบปฏิบัติการประเภทลีนุกซ์ หรือโซลาริส ซึ่งต้องใช้โอเพนซอสตัวอื่นๆ ประกอบด้วย เช่น อาปาเชเว็บเซิร์ฟเวอร์ (Apache Web server) หรือทอมแคท (Tomcat) จาวาคอมไพเลอร์ และโพสต์เกรสเอสคิวแอล (PostgreSQL) ในเรื่องของฮาร์ดแวร์นั้น ดีสเปซต้องการเซิร์ฟเวอร์ที่ค่อนข้างใหญ่ เช่น เอชพี อาร์เอ็กซ์ 2600 (HR rx2600) หรือซันไฟร์ 280อาร์ (SunFire 280R) และมีหน่วยความจำและดิสก์ที่มีความจุเพียงพอ การทำงานของดีสเปซต้องใช้ผู้ดูแลระบบที่มีประสบการณ์สูง

ดีสเปซมีส่วนต่อประสานสำหรับผู้ใช้ทั่วไป ผู้ส่งผลงาน (submitters) และผู้ดูแล (administrators) ส่วนต่อประสานสำหรับผู้ใช้ทั่วไปนั้นใช้สำหรับการเรียกดูรายการและการสืบค้น ผู้ใช้สามารถเรียกดูเนื้อหาได้จากชุมชนหรือสถาบันที่มีสิทธิส่งผลงาน ชื่อเรื่อง ชื่อผู้แต่ง หรือปีพิมพ์ มีการสืบค้นแบบพื้นฐาน (basic search) และการสืบค้นแบบขั้นสูง (advanced search) เพื่อสืบค้นระดับเขตข้อมูลได้อีกด้วย รวมทั้งสามารถแสดงผลการค้นที่เชื่อมโยงไปยังผลงานใหม่ๆ ที่เพิ่งมีการส่งเข้ามา แต่ในการเข้าใช้ในแต่ละรายการถูกจำกัดสำหรับผู้ใช้ที่มีสิทธิในการใช้เท่านั้น ในปี ค.ศ. 2004 ดีสเปซได้ริเริ่มโครงการใหม่โดยร่วมมือกับกูเกิล (Google) เพื่อให้ผู้ใช้กูเกิลสามารถสืบค้นข้ามมาที่คลังเก็บสารสนเทศของดีสเปซได้

เจ้าของผลงานที่ต้องการส่งผลงานต้องลงทะเบียนก่อน เมื่อส่งแฟ้มข้อมูลที่เป็นผลงานเรียบร้อยแล้ว ต้องใส่เมทาดาตาตามเค้าร่างของดับลินคอร์ (Dublin Core metadata schema) ซึ่งประกอบด้วยเมทาดาตาเกี่ยวกับการสงวนรักษา (preservation metadata) และเมทาดาตาประเภทโครงสร้าง (structural metadata) เมทาดาตาที่ผู้ส่งผลงานต้องใส่ คือ ชื่อเจ้าของผลงาน ชื่อเรื่อง ปีพิมพ์ ชื่อชุดและหมายเลขรายงาน ตัวระบุ (identifier) ภาษา หัวเรื่อง/คำสำคัญ สาระสังเขป และผู้สนับสนุน (sponsor) มีเพียงองค์ประกอบ 3 ตัวเท่านั้นที่ต้องการคือ ชื่อผลงาน ภาษา และวันที่ส่งผลงาน ข้อมูลส่วนอื่นๆ ระบบของดีสเปซจะทำให้โดยอัตโนมัติหรือผู้ดูแลเป็นผู้บันทึก

ดีสเปซมีระบบการให้สิทธิแก่กลุ่มผู้ใช้ ผู้ดูแลสามารถเจาะจงผู้มีสิทธิในการส่งผลงาน ผู้ประเมิน (review) ผู้แก้ไข (modify) และผู้บริหารชุมชนหรือทรัพยากรสารสนเทศ ก่อนที่ผลงานจะถูกเก็บลงในคลังเอกสารจะต้องผ่านกระบวนการประเมิน (review process) กระแสงานในดีสเปซมีการประเมินหลายระดับ ผู้ประเมิน (reviewer) สามารถตีคืนผลงานที่พิจารณาเห็นว่าไม่เหมาะสม ผู้รับรอง (approver) เป็นผู้ตรวจสอบกระบวนการส่งผลงาน ในเรื่องของการผิดพลาดในด้านต่างๆ เช่น การลงเมทาดาตา และบรรณธิกรณ์เมทาดาตา (Metadata editor) เป็นผู้มีสิทธิในการแก้ไขเปลี่ยนแปลงเมทาดาตา

โปรแกรมดีสเปซเป็นโปรแกรมที่รับประกันว่าข้อมูลจะไม่มีการเปลี่ยนแปลง แม้ว่าเวลาจะล่วงเลยไปก็ตาม ดีสเปซเก็บรักษาข้อมูลโดยรูปแบบบิตสตรีมของแฟ้มข้อมูล (bitstream format registry) ถ้ารายการที่ส่งมารายการใดไม่อยู่ในรูปแบบที่กำหนดไว้ ผู้ดูแลระบบต้องตัดสินใจว่ารูปแบบนั้นควรถูกขึ้นทะเบียนเอาไว้หรือไม่ ซึ่งรูปแบบที่กำหนดไว้มี 3 แบบ คือ รูปแบบที่ 1 คือ Supported หมายถึง รูปแบบแฟ้มข้อมูลที่มีการประกาศเป็นมาตรฐาน เช่น ทีฟ (TIFF) เอ็กซ์เอ็มแอล (XML) หรือรูปแบบของแฟ้มข้อมูลที่ผู้คิดค้นได้แจกแจงรายละเอียดรูปแบบของแฟ้มข้อมูลให้สาธารณชนรับรู้ เช่น พีดีเอฟ (PDF) รีฟ (RIFF) ซึ่งจะได้รับการเก็บรักษาในระยะยาว รูปแบบที่ 2 คือ Known หมายถึง รูปแบบแฟ้มข้อมูลที่มีการใช้กันในหมู่มาก แต่ผู้คิดค้นไม่ได้ประกาศให้สาธารณชนรับรู้ และรูปแบบที่ 3 คือ Unsupported ถึง รูปแบบของแฟ้มข้อมูลที่ใช้กันน้อย เช่น แฟ้มข้อมูล CAD/CAM หรือแฟ้มข้อมูลที่เป็นโปรแกรม เป็นต้น

ผู้ดูแลระบบสามารถใช้เว็บไซต์ของดีสเปซ ในการหาคู่มือการติดตั้งโปรแกรม การอภิปรายในหัวข้อและกลุ่มเมลลิ่งลิส สถาบันสามารถมีส่วนร่วมในสมาพันธ์ดีสเปซ (DSpace Federation) (http://dspace.org/federation/index.html) ซึ่งเป็นสถานที่ที่ผู้บริหารและผู้ออกแบบระบบมาแสดงความคิดเห็นร่วมกัน

การเปรียบเทียบศักยภาพ

การเก็บถาวรสารสนเทศอิเล็กทรอนิกส์ยังคงอยู่ในระยะแรกเริ่ม แต่อย่างไรก็ตาม มีเครื่องมือสำหรับห้องสมุดขนาดใหญ่และห้องสมุดขนาดเล็กเพื่อดำเนินโครงการการเก็บถาวร โครงการการเก็บถาวรไม่ว่าโครงการใดก็ตามต้องใช้เวลา การวางแผน และความรู้ทางเทคนิค ซึ่งขึ้นอยู่กับว่าห้องสมุดสามารถนำเครื่องมือที่เหมาะสมมาใช้ได้ตรงกับความต้องการและตรงกับทรัพยากรหรือไม่ โครงการล็อคคิสเป็นโครงการที่มีความเป็นไปได้สำหรับห้องสมุดที่ไม่มีเนื้อหาเป็นของตนเองที่จะเก็บในระยะยาว แต่ต้องการมีส่วนร่วมในความพยายามเก็บรักษาผลงานวิทยาศาสตร์ในระยะยาว ประเภทของข้อมูลที่โปรแกรมล็อคคิส สามารถสงวนรักษาได้มีจำกัดมาก เนื่องจากสารสนเทศที่มีการพิมพ์เป็นคาบที่สม่ำเสมอเท่านั้นจึงเหมาะที่จะเก็บด้วยล็อคคิส อย่างไรก็ตาม ยังมีความพยายามที่จะคิดค้นให้ล็อคคิส สามารถเก็บสารสนเทศประเภทอื่นๆ ได้มากกว่าเก็บแค่วารสาร ล็อคคิสเป็นโปรแกรมที่ง่ายและถูกว่าอีพรินทส์และดีสเปซ นอกเหนือไปกว่านี้ล็อคคิสได้เปิดโอกาสในการค้นหาวิธีการแก้ไข เพื่อการแก้ปัญหาของการสงวนรักษาเนื้อหาในระยะยาวผ่านการย้ายและการสำเนาข้อมูล

สถาบันที่ต้องการเก็บรักษานอกเหนือจากวารสารสามารถใช้โปรแกรมอีพรินท์และดีสเปซ เพราะทั้ง 2 โปรแกรมนี้เหมาะสำหรับสถาบันที่ต้องจัดหาช่องทางในการเข้าถึงสารสนเทศที่ผลิตจากหน่วยงานในสถาบันนั้นๆ เอง เพราะฉะนั้นทักษะทางเทคนิคเป็นเรื่องจำเป็นในการติดตั้งทั้ง 2 โปรแกรม โดยเฉพาะอย่างยิ่งโปรแกรมดีสเปซ ซึ่งเก็บสารสนเทศไม่ว่าจะเป็นประเภทใดก็ได้ โปรแกรมอีพรินท์เป็นอีกทางเลือกหนึ่งสำหรับการเก็บรักษาสารสนเทศในเนื้อหาเฉพาะวิชา ขณะที่ดีสเปซมีความเหมาะสมโดยเฉพาะกับสถาบันที่มีขนาดใหญ่ ที่มีความคาดหวังในการเก็บสารสนเทศในระดับใหญ่จากภาควิชาที่มีความหลากหลาย ห้องปฏิบัติการ และชุมชนอื่นๆ ในสถาบัน

สรุป

ด้วยศักยภาพของทั้ง 3 โปรแกรมนี้ ห้องสมุดสถาบันอุดมศึกษาในประเทศไทยน่าจะได้เริ่มมีการศึกษาอย่างจริงจัง เพื่อนำมาใช้ประโยชน์ในการจัดการคลังเก็บเอกสารอิเล็กทรอนิกส์ประเภทวิทยานิพนธ์ของนิสิต นักศึกษา ผลงานวิจัยของอาจารย์ นักวิจัย และบุคลากรของแต่ละสถาบัน รวมทั้งผลงานอื่นๆ ตามความเหมาะสมของแต่ละสถาบันหรือหน่วยงานที่ได้กำหนดนโยบายในการจัดเก็บ การเผยแพร่ รวมทั้งเพื่อการสงวนรักษาเอกสารดิจิทัลในระยะยาว ซึ่งต้องมีการศึกษา การทดลอง การวางแผน และวางนโยบายอย่างเหมาะสมและรัดกุม ทั้งนี้ สถาบันวิทยบริการ จุฬาลงกรณ์มหาวิทยาลัย กำลังศึกษาโปรแกรมดีสเปซ เพื่อนำมาทดลองใช้ในการจัดการคลังเก็บวิทยานิพนธ์อิเล็กทรอนิกส์ และจะขยายผลต่อไปถึงผลงานต่างๆ ของประชาคมจุฬาลงกรณ์มหาวิทยาลัย