

# 雲端運算最新發展： 建構數位圖書館於雲端中

胡毓忠

新趨勢網路實驗室

國立政治大學資訊科學系

台北市文山區指南路二段64號 11605

hu@cs.nccu.edu.tw

# 研究動機和目的

- 找出數位圖書館建置於雲端中的正負面因素
- 勾勒出數位圖書館於雲端中所將面臨的挑戰
- 雲端式數位圖書館的現況與研究
- 圖書資訊系統與數位內容放置於雲端中
- Web 3.0 Linked Library Data (LLD)
- 整合著作權與隱私權兩者的保護規範

# 雲端運算環境的介紹

# 雲端運算環境的特色

- 是新思維還是舊瓶裝新酒？
- 透過SLA的傳統式委外和雲端式委外對比
- Utility Computing資源使用模式
- 雲端運算的服務模式
  - 架構式服務 (IaaS)
  - 平台式服務 (PaaS)
  - 軟體式服務(SaaS)
- 部署方式：私有雲、公共雲、社群雲、混和雲
- 雲端式數位圖書館的服務模式與部屬方式？

# 雲端運算環境的現況

- 雲端運算環境和Web環境的整合
  - Web 1.0 (Original Web)：關連式資料庫(RDB)
  - Web 2.0 (Social Web)：經驗分享和推薦
  - Web 3.0 (Semantic Web)：串連式開放資料 (LOD)
- Web Services和雲端環境的結合
- 已有多家知名企業提供雲端環境
- 使用者已經在不自知的情況下使用
- 政府機構也在考慮建構公共雲
- 開放式雲端環境架構：Hadoop, NEBULA

# Hadoop是什麼？

What Is Hadoop ?



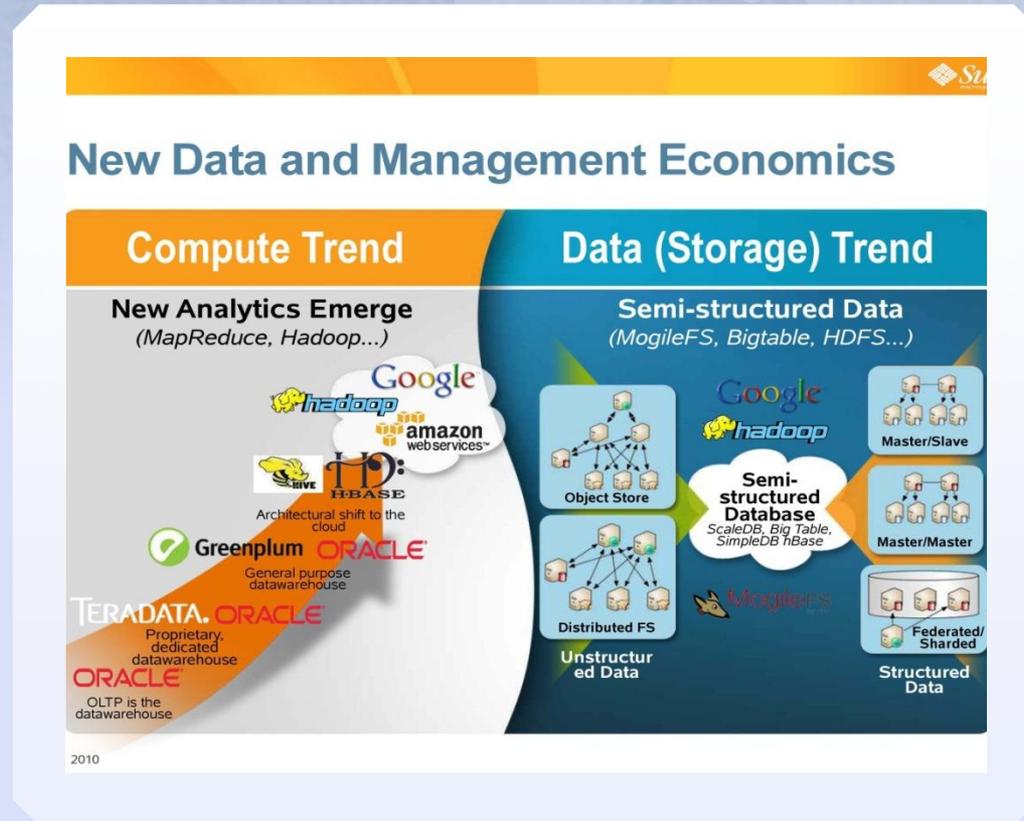
*“Flexible and available architecture for large scale computation and data processing on a network of commodity hardware”*

Open Source + Hardware Commodity = IT Costs Reduction

2010

Source: Hadoop Architecture, Philippe Julio – Sun Microsystems, France

# Hadoop 雲端架構



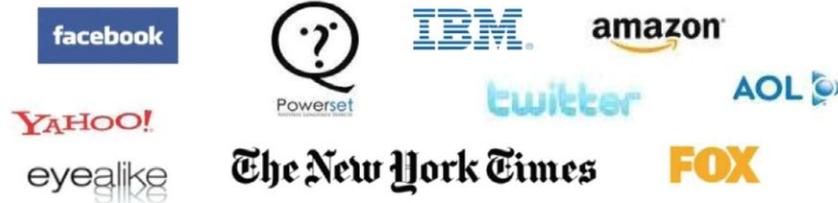
Source: Hadoop Architecture, Philippe Julio – Sun Microsystems, France

# 有誰在用Hadoop?

## Who Used Hadoop ?



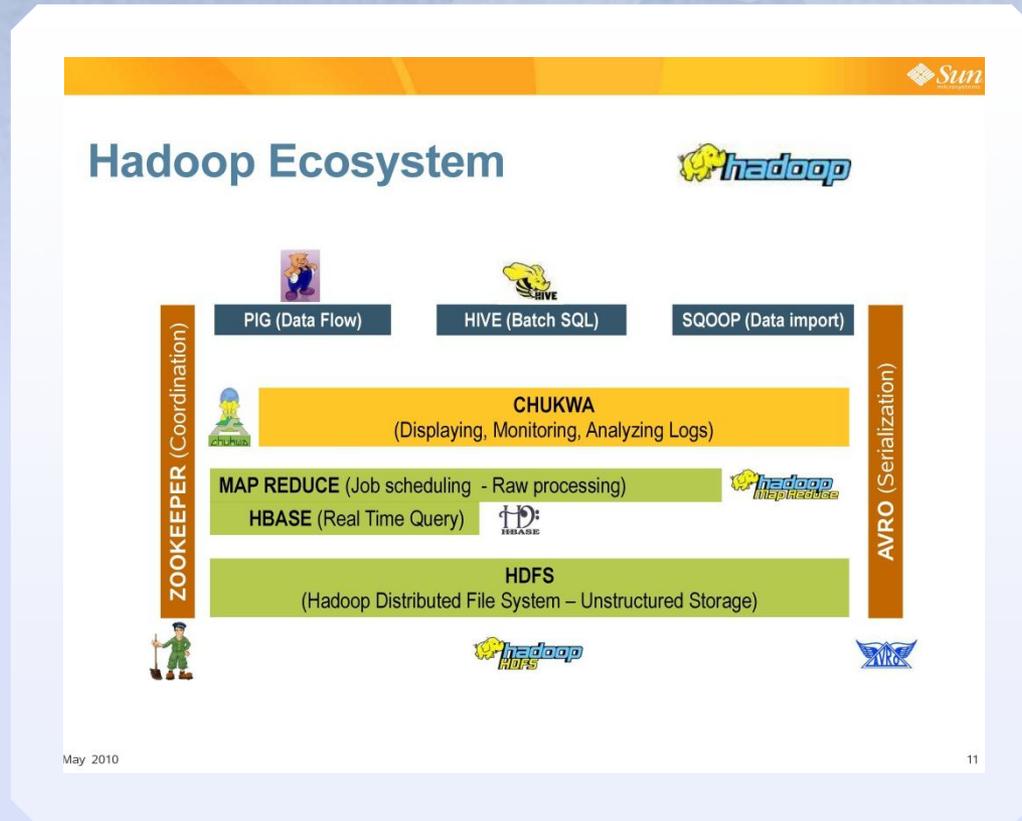
Top level Apache Foundation project   
Large, active user base, mailing lists, user groups  
Very active development, strong development team



<http://wiki.apache.org/hadoop/PoweredBy>

Source: Hadoop Architecture, Philippe Julio – Sun Microsystems, France

# Hadoop Ecosystem



Source: Hadoop Architecture, Philippe Julio – Sun Microsystems, France

# 建置數位圖書館於雲端中

# 數位圖書館建置於雲端的考量

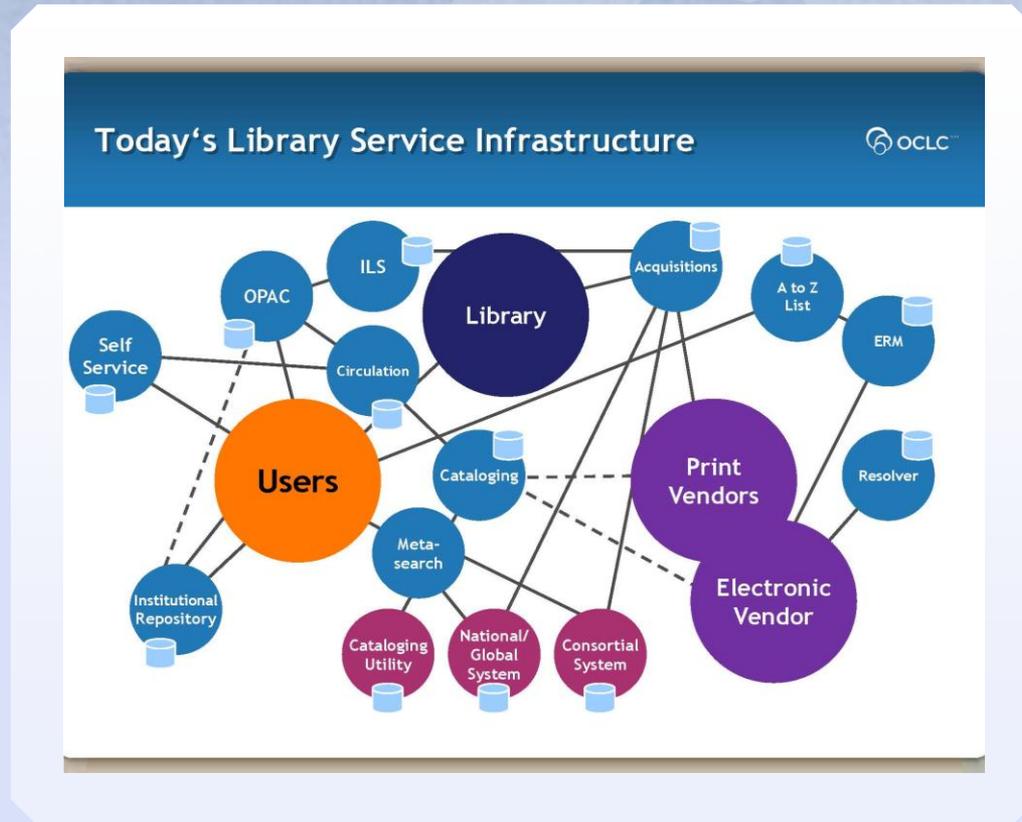
## 正面考量

- 符合時代潮流趨勢
- 節省建置與開發成本
- 容易進行數位圖書館際整合和資源的分享
- 數位圖書館其數位內容與管理系統的可擴充性
- 更容易和雲端化的數位內容出版業者進行系統與內容的整合

## 負面考量

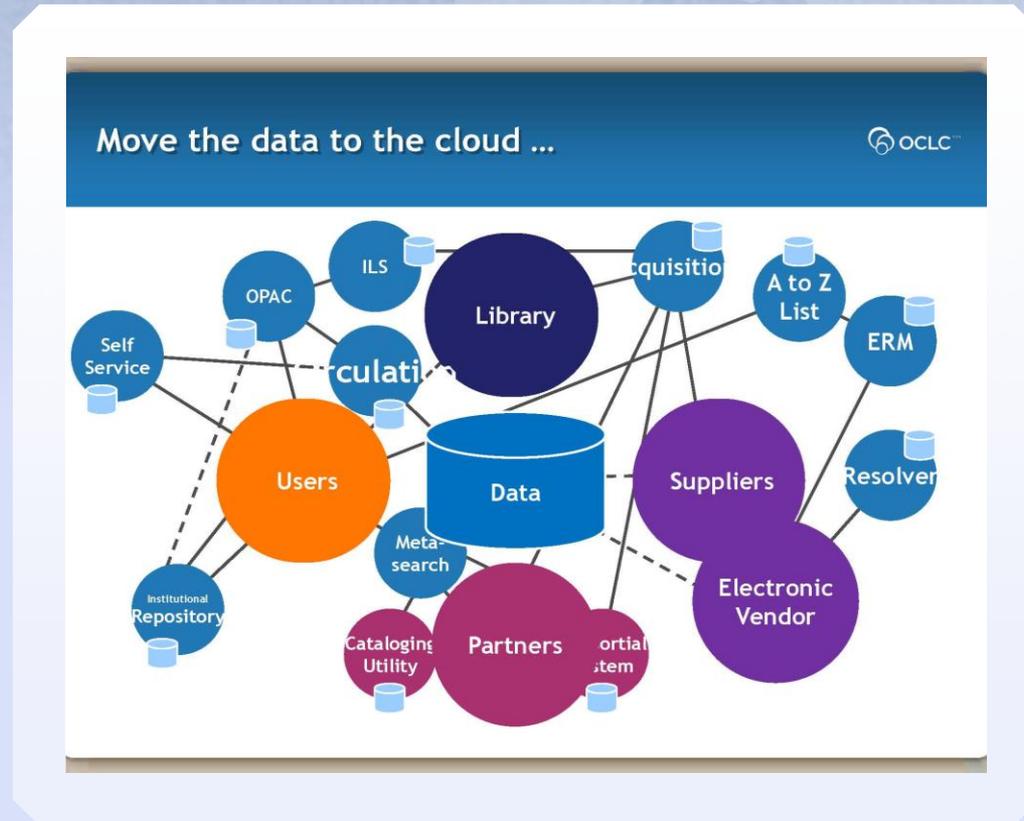
- 時機尚未成熟（含商業模式的不確定性）
- 雲端式委外的風險
- 現有圖書資訊系統和內容轉換到雲端上的困難
- 尚未有完整的雲端式數位圖書系統平台可供選擇
- 著作權與隱私權保護責任歸屬的不確定性
- 無法整合到尚未到來的雲端化大學資訊系統

# 現有圖書資訊服務系統架構



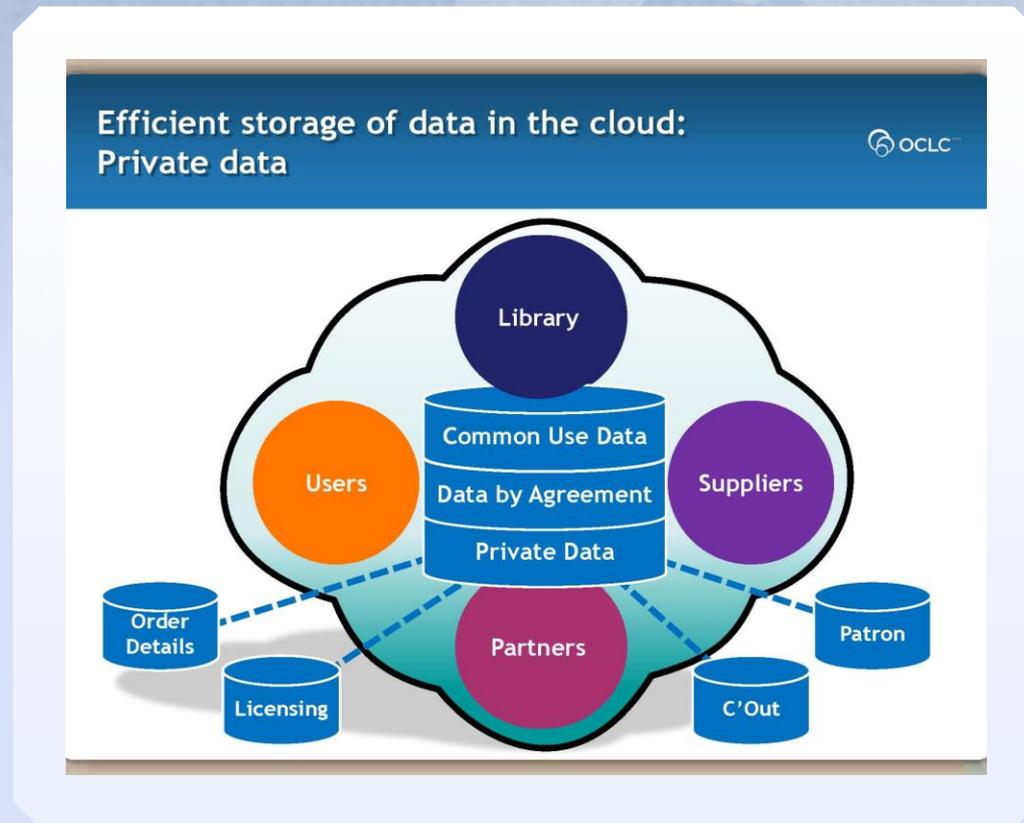
Source: Weinberger, N. (2010), Weinberger N. "The Future is Web-Scale", IFLA 2010

# 雲端式圖書資訊服務系統架構



Source: Weinberger, N. (2010), Weinberger N. "The Future is Web-Scale", IFLA 2010

# 資料於雲端式圖書資訊服務系統



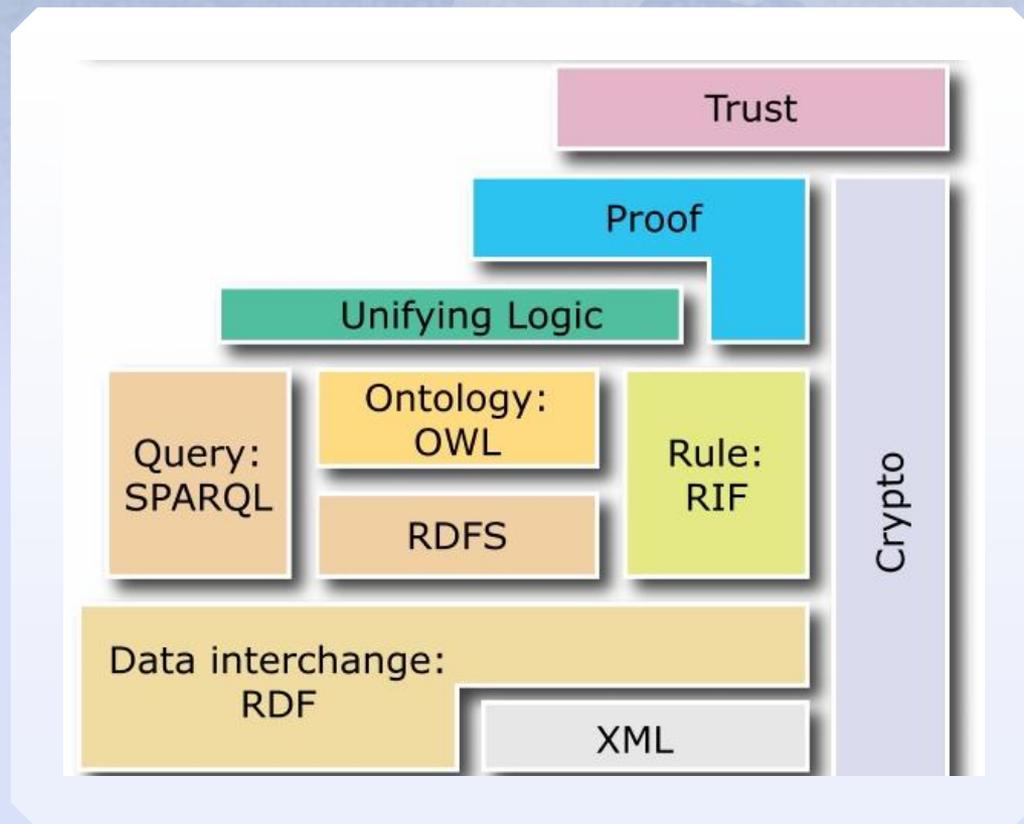
Source: Weinberger, N. (2010), Weinberger N. "The Future is Web-Scale", IFLA 2010

# 數位圖書館雲端化將面臨的挑戰

- 委外（含完全與部分）或不委外於雲端
- 現有圖書資訊系統與內容的轉換與改變
- （商業）服務模式的不確定性
- 數位內容與系統所提供的流通、分享、使用與保護
- 提供各種數位內容不限時間和地點的使用
- 數位內容著作權保護（含合理使用）
- 使用者個人資料與隱私權保護
- 整合著作權保護與隱私權保護規範
- 與現有或未來大學資訊系統的整合

# 雲端式數位圖書館的研究

# 語意網(Web 3.0)的架構示意圖



Source: W3C --<http://www.w3.org/2007/03/layerCake.svg>



# Linked Library Data Reference Sites

## Let's start with references

### *Landmark Linked Data implementations by library actors*

- Swedish National Library's Libris catalogue and thesaurus [libris.kb.se/](http://libris.kb.se/)
- Library of Congress' vocabularies, including LCSH [id.loc.gov/](http://id.loc.gov/)
- DNB's Gemeinsame Normdatei [d-nb.info/gnd/](http://d-nb.info/gnd/)
- BnF's RAMEAU subject headings [stitch.cs.vu.nl/rameau/](http://stitch.cs.vu.nl/rameau/)
- OCLC's DDC classification and VIAF [dewey.info/](http://dewey.info/) [viaf.org/](http://viaf.org/)
- NL of Hungary's catalogue and thesauri [oszkdk.oszk.hu/resource/DRJ/404](http://oszkdk.oszk.hu/resource/DRJ/404)

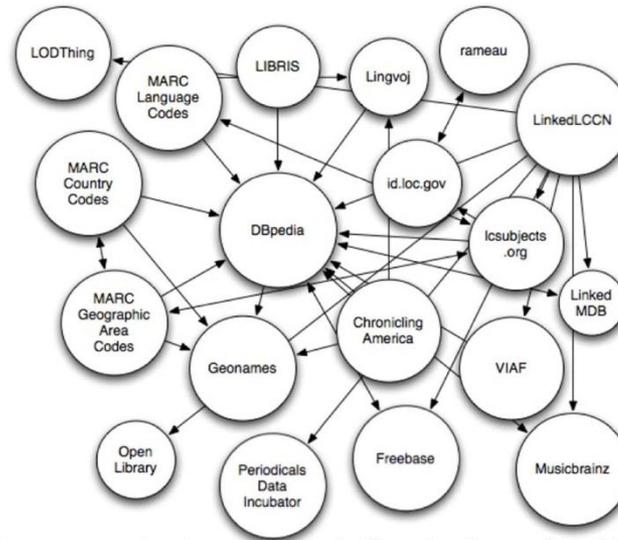
### *Also relevant!*

- STW economy thesaurus [zbw.eu/stw](http://zbw.eu/stw)
- Social Science thesaurus [lod.gesis.org](http://lod.gesis.org)
- GEMET environmental thesaurus [eionet.europa.eu/gemet](http://eionet.europa.eu/gemet)
- Agrovoc [aims.fao.org/](http://aims.fao.org/)
- New York Times subject headings [data.nytimes.com/](http://data.nytimes.com/)
- Scientific publications (among others) [dblp.rkbexplorer.com/](http://dblp.rkbexplorer.com/)

Source: W3C Library Linked Data Incubator Group, Antoine Issace, Vrije Universiteit Amsterdam  
Talis Open Day: Linked Data and Libraries, London, July 21st 2010

# The Linked Library Data Cloud

Linked Library Cloud mid-2010



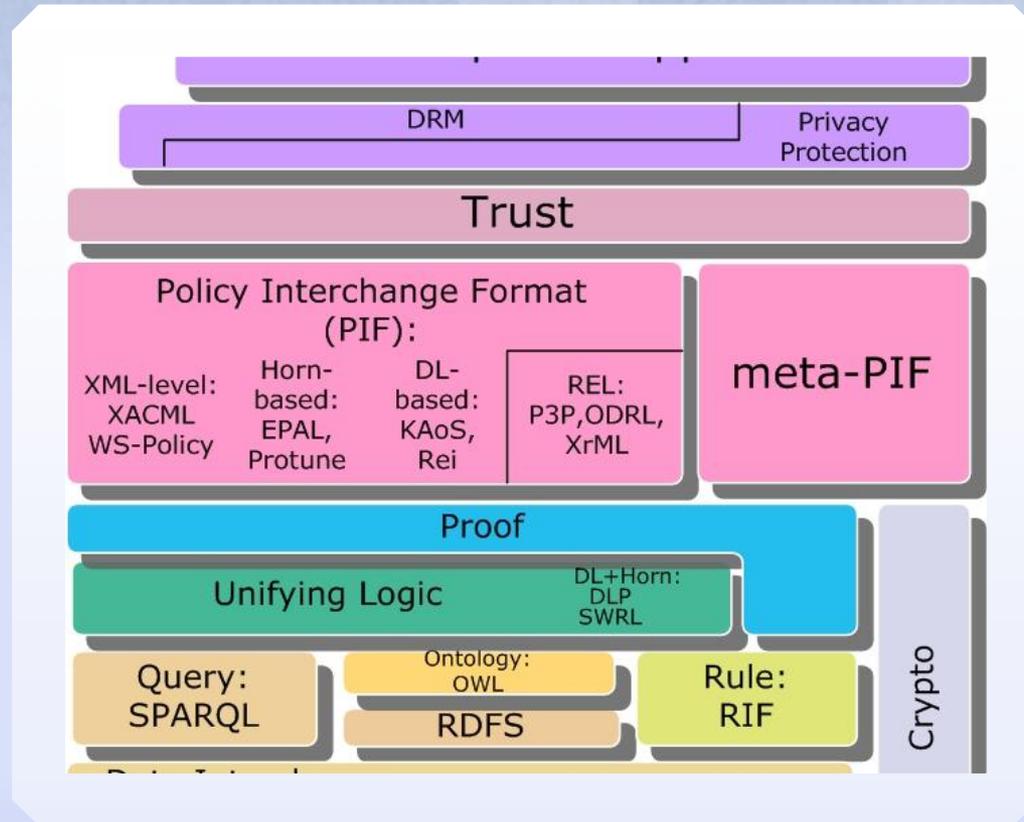
Plus:

- Germany NL
- Hungary NL
- STW
- GEMET
- NYT
- Agrovoc

[Ross Singer, Code4Lib2010] <http://code4lib.org/conference/2010/singer>

Source: W3C Library Linked Data Incubator Group, Antoine Issace, Vrije Universiteit Amsterdam  
Talis Open Day: Linked Data and Libraries, London, July 21st 2010

# SemPIF電腦語言式規範架構圖



Source: Hu, Y.J. and H. Boley, "SemPIF: A Semantic Meta-Policy Interchange Format for Multiple Web Policies", 2010 IEEE Web Intelligence (WI) Conference Aug. 31-Sep. 3, 2010

# A Scenario of Digital Library Subscription

## Server Side:

The NCCU university library has subscribed to IEEE, ACM, Springer, and Google Books digital library services in the cloud, which provide a set of eJournal, eBooks, and other multimedia contents' access rights for its authorized students and staff members.

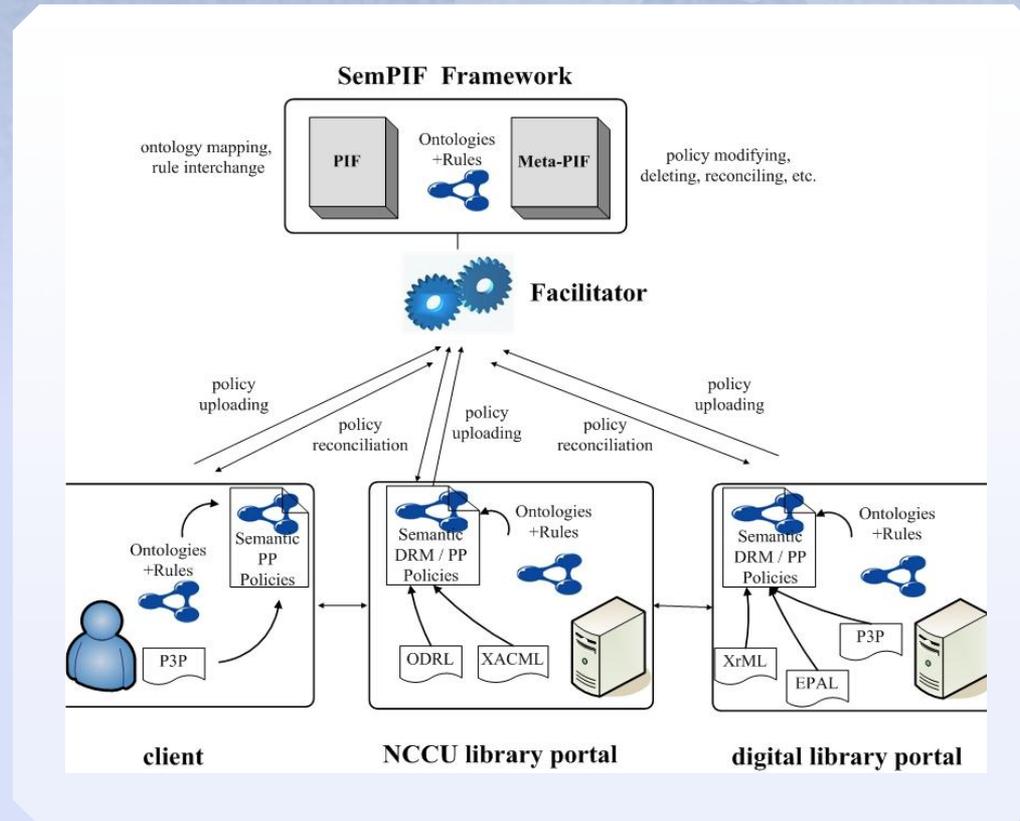
There are two types of policy for an NCCU and IEEE Web servers:

- one for DRM of content rights
- the other for the declaration of privacy statements

## Client Side:

A student, John, as a Web client has privacy protection policies to address how and what of his personal data can (or cannot) be collected, retained, or disclosed from possible Web servers in the cloud.

# 整合著作權與隱私權保護規範



Source: Hu, Y.J. and H. Boley, "SemPIF: A Semantic Meta-Policy Interchange Format for Multiple Web Policies", 2010 IEEE Web Intelligence (WI) Conference Aug. 31-Sep. 3, 2010

# 結論和未來展望

- ❑ 圖書資訊系統雲端化的管理平台提供已經略見端倪，請區隔你要的是圖書資訊系統或（且）是數位內容放到雲端上
- ❑ Web 3.0的Linked Library Data (LLD)將上傳到Web數位圖書資訊系統的Context (Metadata)與其管理Content串連已略具初步成果，可考慮未來是否將LDD放到雲端上
- ❑ 數位圖書館全面雲端化的時機是否已經成熟需要再觀察，請思考所將面臨的挑戰和（商業）服務模式的改變
- ❑ 整合資訊科學和圖書資訊學來釐清上述議題讓台灣圖書資訊系統可以和世界接軌
- ❑ 數位內容的著作權保護和使用個人資料保護將會是數位圖書資訊系統雲端化後的另一大挑戰