

The Challenges of Building a Digital Preservation System

July 2011



Ido Peled

Ex Libris Rosetta Product Manager

Copyright Statement

All of the information and material inclusive of text, images, logos, product names is either the property of, or used with permission by Ex Libris Ltd. The information may not be distributed, modified, displayed, reproduced – in whole or in part – without the prior written permission of Ex Libris Ltd.

TRADEMARKS

Ex Libris, the Ex Libris logo, Aleph, SFX, SFXIT, MetaLib, DigiTool, Verde, Primo, Voyager, MetaSearch, MetaIndex and other Ex Libris products and services referenced herein are trademarks of Ex Libris, and may be registered in certain jurisdictions. All other product names, company names, marks and logos referenced may be trademarks of their respective owners.

DISCLAIMER

The information contained in this document is compiled from various sources and provided on an "AS IS" basis for general information purposes only without any representations, conditions or warranties whether express or implied, including any implied warranties of satisfactory quality, completeness, accuracy or fitness for a particular purpose.

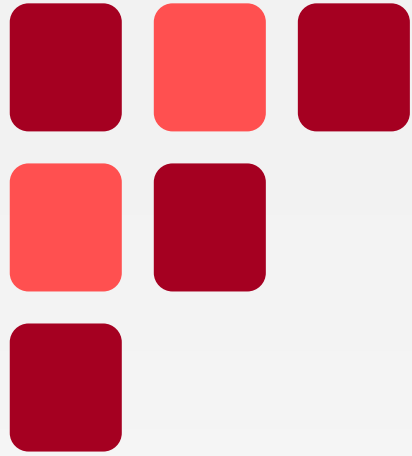
Ex Libris, its subsidiaries and related corporations ("Ex Libris Group") disclaim any and all liability for all use of this information, including losses, damages, claims or expenses any person may incur as a result of the use of this information, even if advised of the possibility of such loss or damage.

© Ex Libris Ltd., 2010

The Case for Digital Preservation

What does *Google Videos* and *Blueberry Muffin* have in common ?





Why Digital Preservation



What is at Stake ?

“Information is the oxygen of the modern age. It seeps through the walls topped by barbed wire, it wafts across the electrified borders”

Ronald Reagan

The collage consists of several overlapping web pages:

- The New York Times:** Front page dated Tuesday, May 8, 2012. Headlines include "Big G.O.P. Donors Adopt Wait-and-See 2012 Tack" and "The vaulted Republican network of high-dollar donors and fund-raiser has been slow to support Mitt Romney".
- CNN:** News site with a headline "Bin Laden's young bride" and a photo of a woman in a headscarf.
- ESPN:** SportsCenter website with a "SO HIGHLIGHT OF THE NIGHT" featuring a basketball game.
- Googleshare Map:** A world map showing search volume by country. A table below the map provides the following data:

Country name	Country code	Googleshare	PageCount for country name	keyword
China	CN	0.09903181158	72000000	718000
India	IN	0.14123006832	43900000	620000
USA	US	0.38111111111	117000000	719000
Indonesia	ID	0.18514261426	30300000	981000
Brazil	BR	0.23486666666	22500000	528000
Pakistan	PK	0.196875	18000000	251000
Bangladesh	BD	0.2114009959	9840000	189000
Nigeria	NG	0.26763486477	9640000	258000
Russia	RU	0.27291666666	19000000	524000
Japan	JP	0.10471124821	69800000	689000
Germany	DE	0.10420254431	53700000	993000
UK	UK	0.09229084511	16900000	779000
France	FR	0.07201754388	11400000	821000
Canada	CA	0.0629411761	10200000	619000
Korea	KR	0.23644837791	23800000	599000
Poland	PL	0.15324074071	21600000	331000
Hungary	HU	0.15075767671	13000000	199000
Iraq	IQ	0.20802259881	17200000	461000
Greece	GR	0.20618430391	14600000	301000
Mexico	MX	0.08629240952	69000000	642000
Ulyss	LY	0.20230607968	47700000	96500
- Email:** From the Rosetta Advisory Workgroup, dated 10:02 08/02/2011. The email discusses consolidating Rosetta Advisory Workgroups and lists:
 1. Format Library Workgroup chaired by the National Library of New Zealand and Catherine Holzem
 2. Consolidated Rosetta Advisory Workgroup (as I have to ask for a checkbox from the systems and I'll be co-chairing the Workgroup from the F.L. Lib. side)

Who Should Be Concerned ?

“A popular government without popular information or the means of acquiring it, is but a prologue to a farce, or a tragedy, or perhaps both”

James Madison

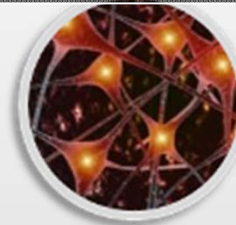
International
Libraries



University Library



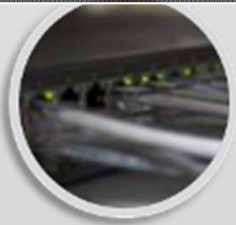
Researchers



Cultural Heritage
Organizations



Media Companies



State & Local
Governments



Federal Agencies



Specialist Archives

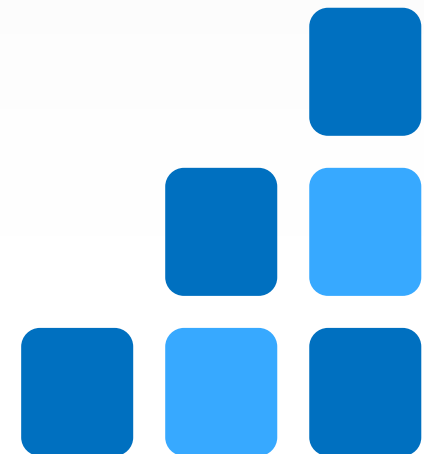


Challenges of Digital Preservation





Building a Digital Preservation System



Challenge I – Digital Preservation Definition



<http://bcservice.net/2011/01/network-disaster-recovery-pt-3/>

Challenge I – Digital Preservation Definition

“Digital preservation combines **policies, strategies** and **actions** to **ensure access** to **reformatted and born digital** content regardless of the challenges of **media failure** and **technological change**.

The goal of digital preservation is the **accurate rendering of authenticated content over time**”

ALA (American Library Association)



Challenge I – Digital Preservation Definition

• Digital preservation is the management of digital information **over time**. Preservation of digital information is widely considered to require more constant and ongoing attention than preservation of other media. ...

en.wikipedia.org/wiki/Digital_preservation

• The **long-term** maintenance and upgrade of digital files on digital storage medium

hosting.collectionsaustralia.net/capture/course/glossary.html

• The process of **ensuring** that a digital object is **accessible** over the **long term**.

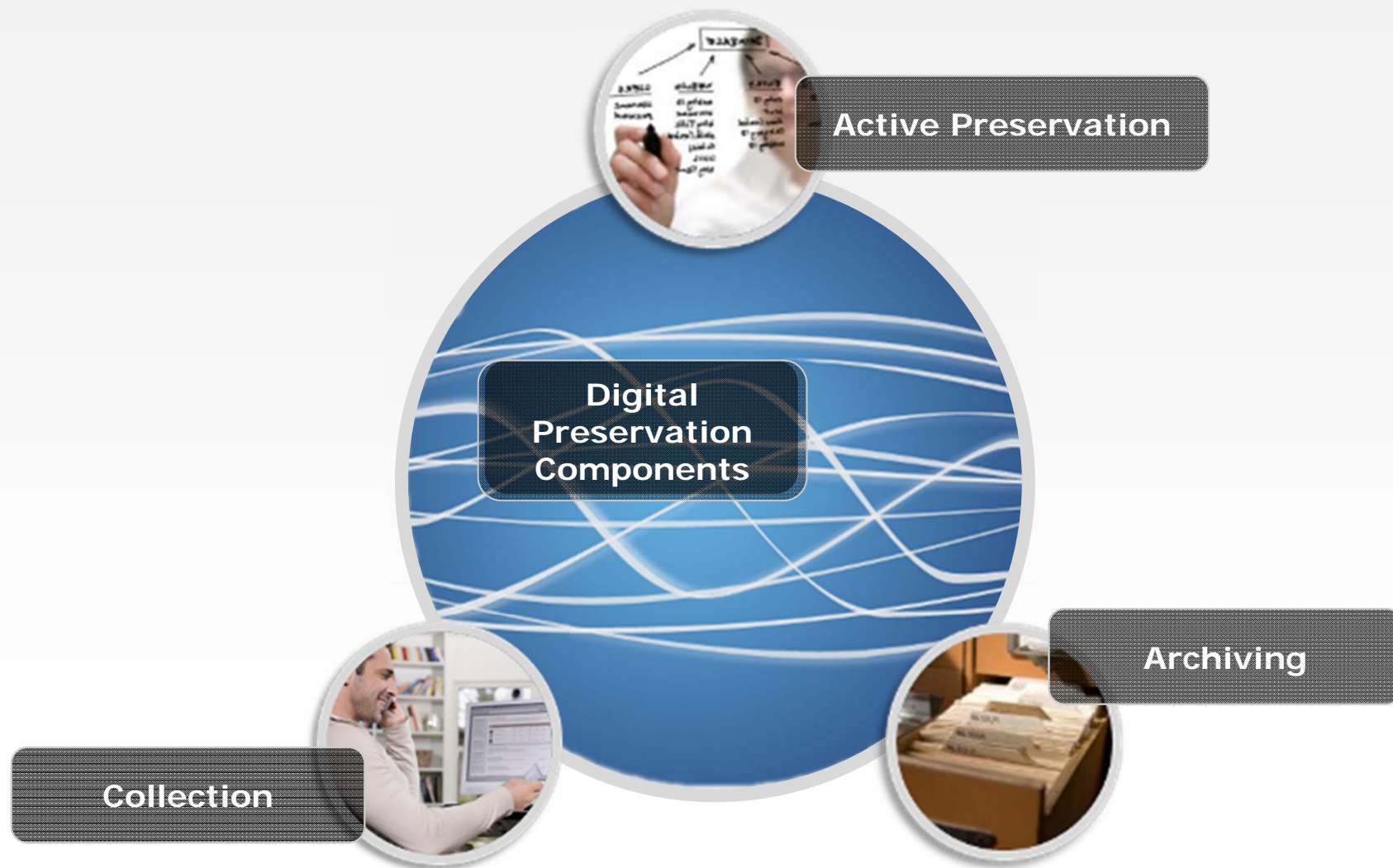
library.manoa.hawaii.edu/scholarspace/glossary.html

• Refers to the series of managed **activities** necessary to **ensure continued access** to digital materials for as long as necessary.

www.digitaltraditions.net/html/Glossary.cfm



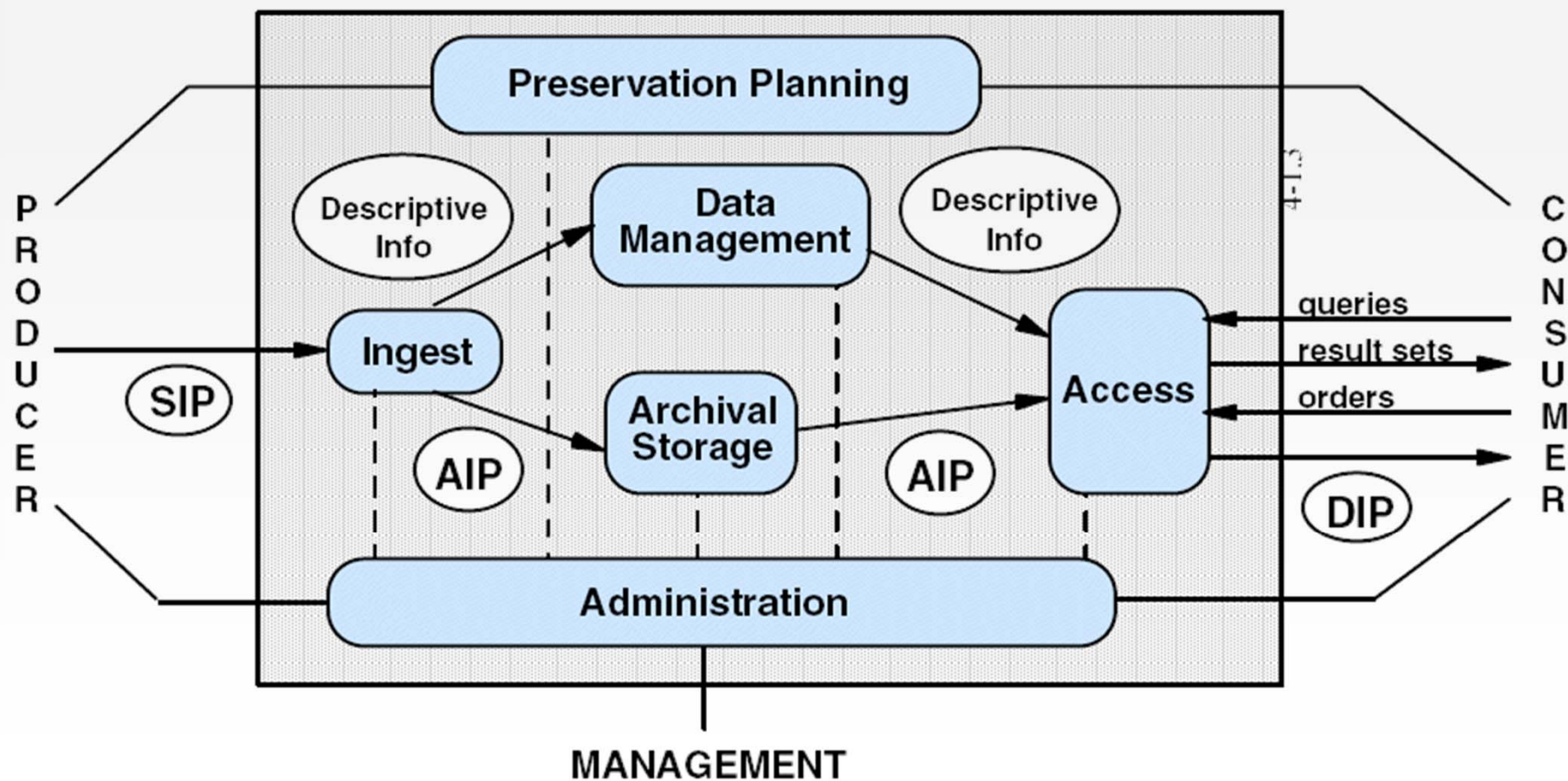
Challenge I – Digital Preservation Definition



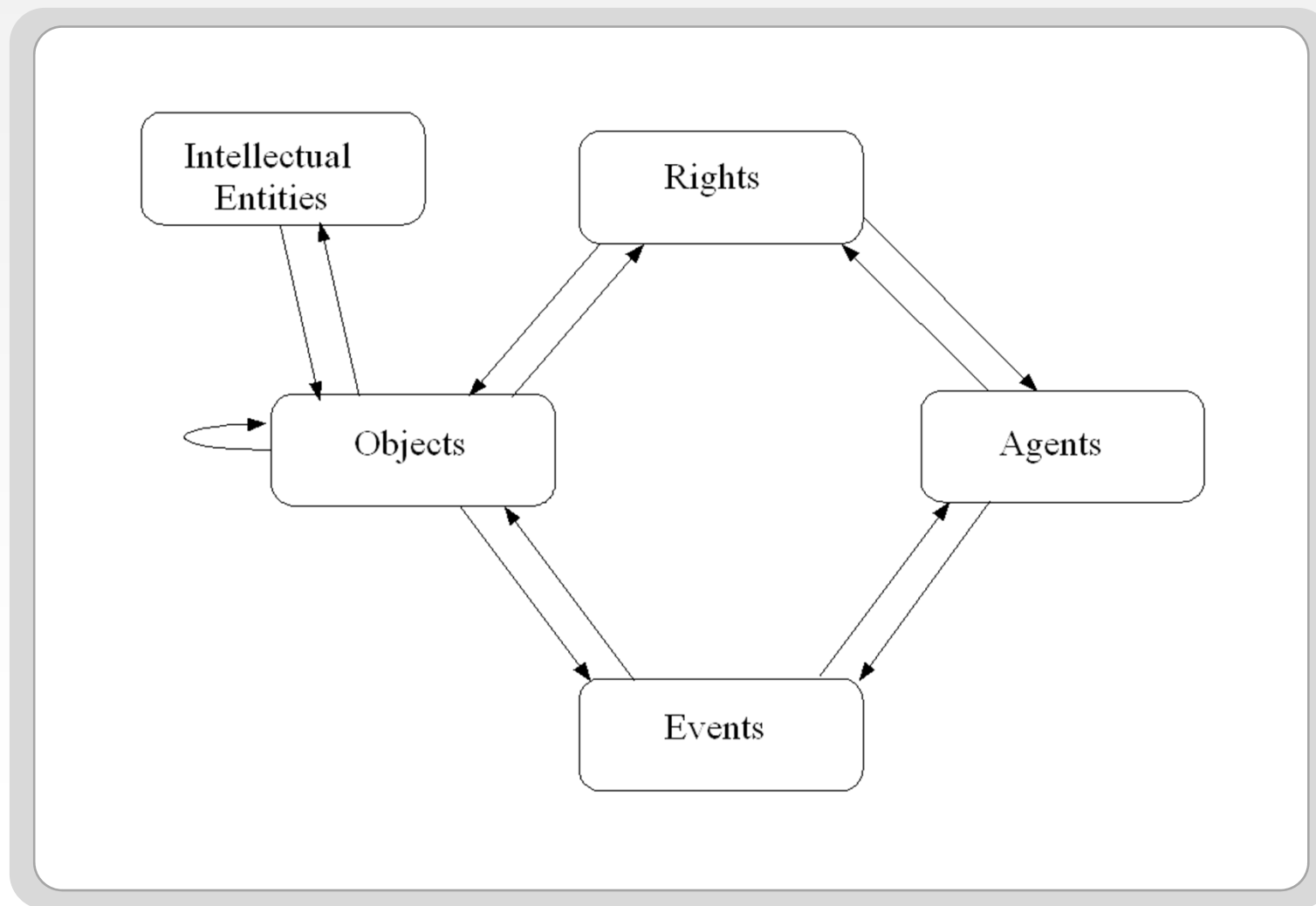
Challenge II – Data Model

- Support structure and metadata for long term digital preservation
- Variety of formats
- Variety of institution types
- Conform to standards and best practices

Challenge II – Data Model



Challenge II – Data Model



Challenge III – Characterization



<http://www.flickr.com/photos/sreejithk2000/2385193167/>

Challenge III – Characterization

- A framework around the different tools including:
 - A technical analyst workbench to handle the different issues
 - Rule based decision making to automate the process
 - Use the local format library to define rules of what to use on each format

Challenge IV – Existing Infrastructure

- Institutions have infrastructure in place
- Not creating yet another system
- Integrate with existing environment.

Challenge IV – Existing Infrastructure

Validation
Plug-In

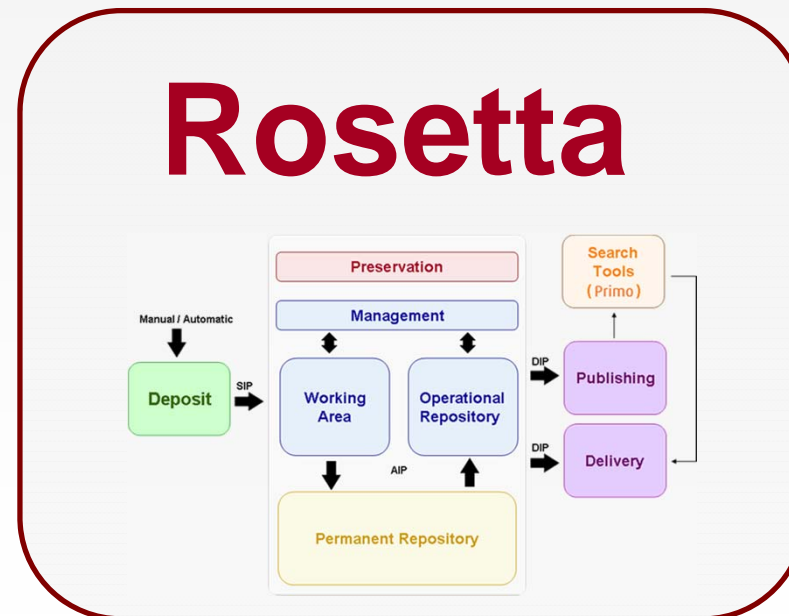
Enrichment
Plug-In

Migration
Plug-In

Deposit API

IE Parsing Tools

Authentication



Viewers

UI Customization

SRU/SRW

OAI

Deep Links

Challenge V – The System as a Bottleneck



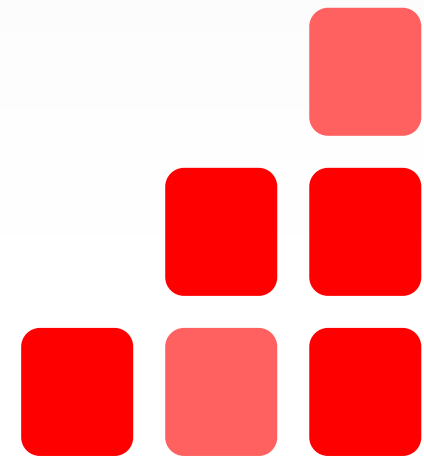
<http://www.psdgraphics.com/psd/green-exit-sign-psd/>

Challenge V – The System as a Bottleneck

- All preservation worthy information is stored on disk
- Storage abstraction later
- Storage is open and based on standards
- Information can be restored into Rosetta or into another system

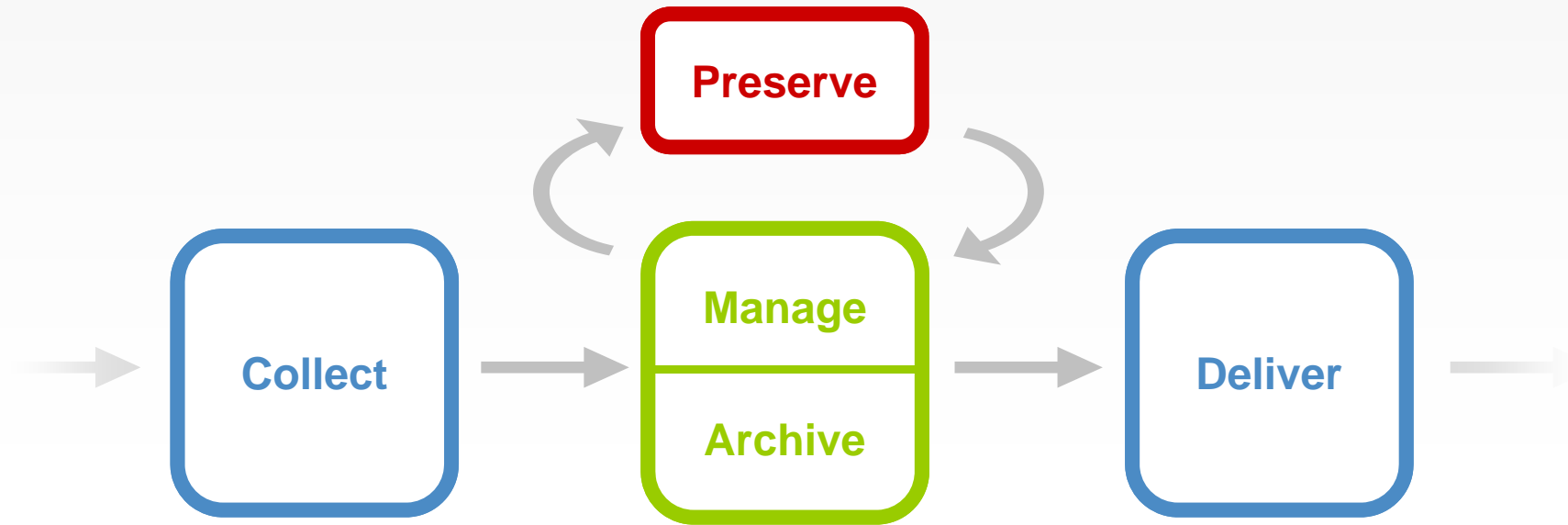


Ex Libris Rosetta

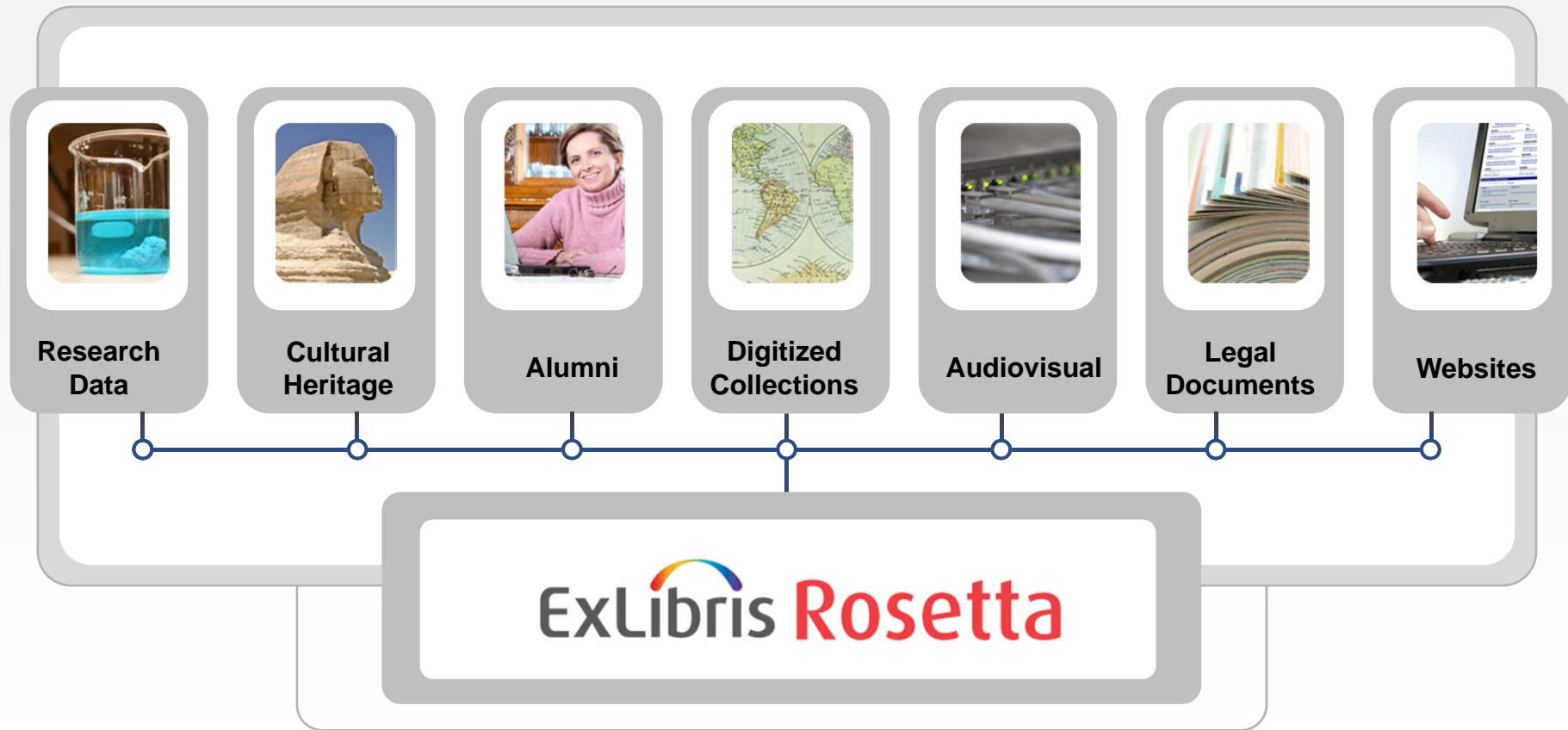


Ex Libris Rosetta

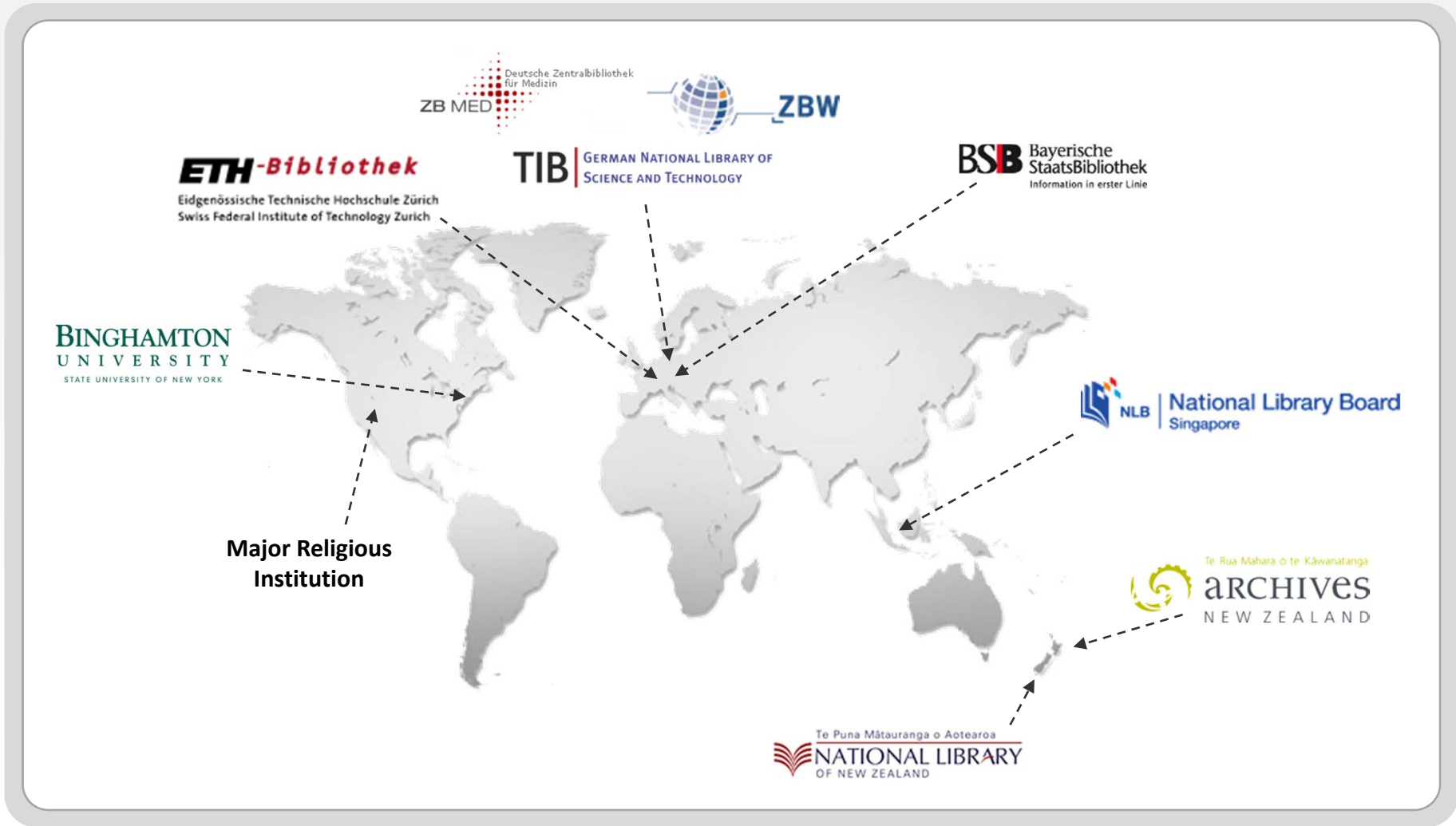
Ex Libris Rosetta helps institutions in **collecting, managing, archiving** and **preserving** their digital collections, ensuring its **data integrity** and **access over time**



What Can Rosetta Do For You ?



Our Customers Around the World



Conclusions

- Digital materials are easily created and on a large scale
- Awareness for digital preservation is growing
- Challenges are well defined

Rosetta

- These challenges represent a small proportion of the requirements and objectives of the Rosetta digital preservation solution
- Complete preservation solution
- Scalable, flexible and extensible

Readings

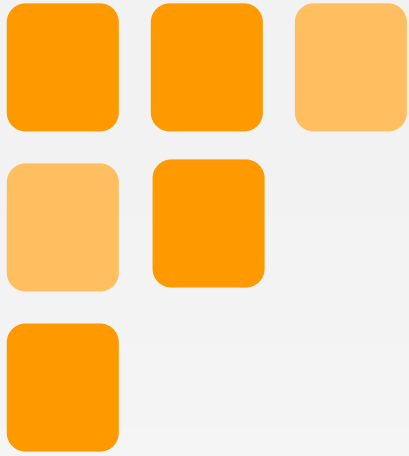
Reference Model for an Open Archival Information System (2002). Available at: <http://public.ccsds.org/publications/archive/650x0b1.pdf>

PREMIS, Data Dictionary for Preservation Metadata(2008),
Available at: <http://www.loc.gov/standards/premis/v2/premis-2-0.pdf>

100 Year Archive Requirements Survey (2008),
Available at:
http://www.snia.org/forums/dmf/knowledge/100YrATF_Archive-Requirements-Survey_20070619.pdf

Sustainable Economics for a Digital Preservation, BRTF (2010),
Available at: http://brtf.sdsc.edu/biblio/BRTF_Final_Report.pdf

Mind the Gap (2006),
Available at:
http://www.dpconline.org/index.php?option=com_docman&task=doc_download&gid=340



Thank You !!

Ido.Peled@ExLibrisGroup.com

