Old silos, new silos, no silos

From redundancy to aggregation or distribution?

Lukas Koster
Library Systems Coordinator
Library of the University of Amsterdam
@lukask
l.koster@uva.nl
SWIB12, Köln, November 26-28, 2012

http://www.flickr.com/photos/alpoma/4786094569

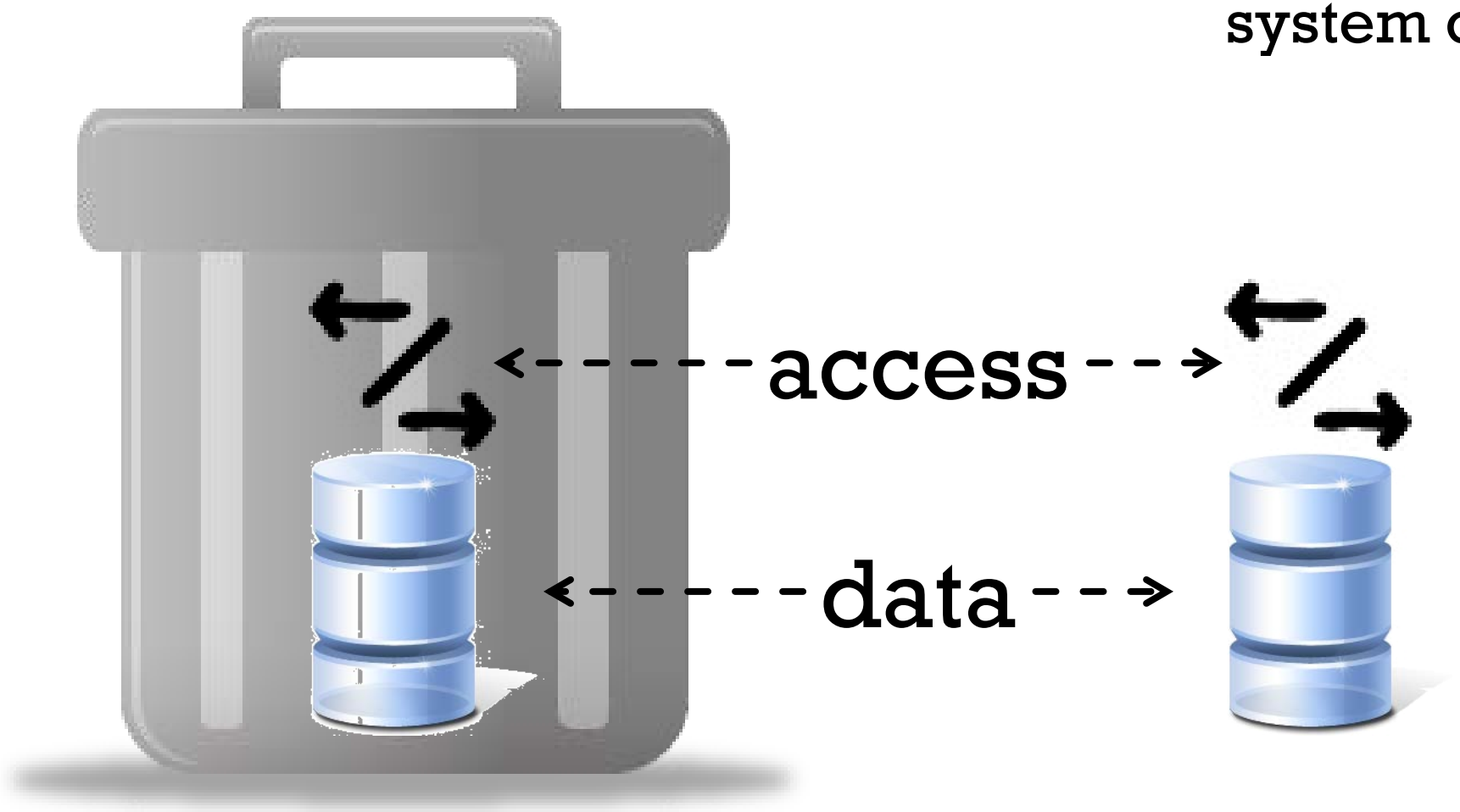
What is a silo?

"System incapable of reciprocal operation with other, related information systems"

http://en.wikipedia.org/wiki/Information_silo

http://www.flickr.com/photos/7716310@N03/1472185959

"System containing data, only accessible via internal system calls"



SILO **NO SILO**

Application Programming Interface

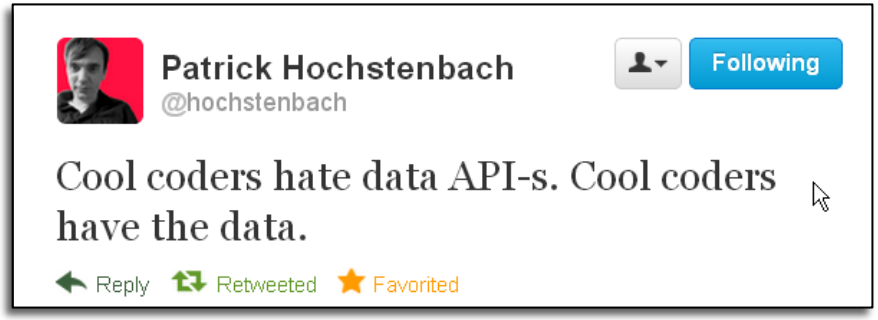
API

Direct access

system call

data call

(LOD)



http://twitter.com/hochstenbach/status/268033173366124544

http://twitter.com/jindrichmynarz/status/269089964103434242

Focus

Traditional library content: publications

http://commons.wikimedia.org/wiki/File:Latin_dictionary.jpg

Story

Issue 2

November 2005

SMUG 4 EU

EUROPEAN NEWSLETTER FOR SFX/METALIB USERS

Andrea Kroneisl Beate Rusch Damyanti Patel Duccio di Blasi Lukas Koster Repke de Vries

In this issue:

- In Focus: Knowledge bases
- Interviewed: Chris Roberts
- In Portrait: Portugal
- Inside: Pivotal
- In London: Eye on SMUG

On death, Google and karaoke

September 2005, a crowded pub in London's Bloomsbury area, could be The Lamb, The Swan or The Frog. Lager, Bitter and Guinness are running continuously out of shiny golden taps. The later the evening, the more rosy faces around: among them a group with white name cards awry clipped at their collars. They seem to have just come in from a conference centre nearby.

A (*loud*): Have you heard? It's dead. They all say it's dead.
B (*startled*): Who's dead?
A (*laughing*): I guess, there won't even be a funeral. Nobody seems to be really sad.
C (*sighing*): Google, Google, ...
B: How many glasses have you had? What are you talking about?
C and A (*singing in a kind of canon*): Google, Google, Google.
A: Seriously, don't you agree, that federated search systems are dead? We can't compete with them, can we?
B: But we just started with MetaLib and SFX, statistics look promising ... Especially after we invested in trainings, special BA courses, introducing MetaLib and SFX. I could show you the numbers.
C (*sighing again*): You need courses, that's for sure.
A: But you won't get those kids back that you have already lost. They will stay with Google ...
B: Please, don't sing.
A (*a little peeved*): Don't you like my voice? (*happy again*) We could go to a karaoke bar afterwards.
B: But what is the alternative? Build big databases like Google does? Harvest data till you drop? (*A pause. Then thinking loud*) Probably we do need both. Big databases as well as federated search. You will never be able to harvest all the worlds data.
C (*mumbling*): Perhaps only a limited number of distributed harvested indexes around the world ...
A: Maybe I mixed it up: not federated search is dead, but the dream, that federated search is the ultimate solution.
C (*again awake*) : Yes! That must be it! And then federated search can survive and be used for accessing those indexes! *The bell rings for last orders.*
C: Have another beer. It's my round.
A + B: Yeess ... And then the karaoke bar ... (*singing*) Yesterday, all my troubles seemed so far away ...

EAVESDROPPED BY LUKAS KOSTER AND BEATE RUSCH

CONTACT:
LUKAS.KOSTER(@)KB.NL
RUSCH(@)ZIB.DE

Issue No 2 was edited by

Beate Rusch, KOBV Berlin
Damyanti Patel,
Royal Holloway London
Andrea Kroneisl,
The Austrian Library Network and Service LTD Vienna
Contact: editors(@)smug-4-eu.org

Story

EAVESDROPPED BY LUKAS KOSTER AND BEATE RUSCH

CONTACT:
LUKAS.KOSTER(@)KB.NL
RUSCH(@)ZIB.DE

http://igelu.org/wp-content/uploads/2010/10/smug4eu_issue2.pdf

Issue 2

November 2005 !

7 years ago!

A: *Federated search systems are dead*
But you won't get those kids back that you have already lost. They will stay with Google

B: *But what is the alternative? Build big databases like Google does?*
Harvest data till you drop?
Probably we do need both. Big databases as well as federated search.

C: *Perhaps only a limited number of distributed harvested indexes around the world.*
Federated search can survive and be used for accessing those indexes!

Issue 2

November 2005 !

7 years ago!

A: *Federated search systems are dead*
But you won't get those kids back that you have already lost. They will stay with Google

B: *But what is the alternative? Build big databases like Google does?*
Harvest data *till you drop?*
Probably we do need both. ***Big data****bases as well as federated search.*

C: *Perhaps only a limited number of* ***distributed harvested indexes*** *around the world.*
Federated search can survive and be used for accessing those indexes!

Issue 4

February 2007

SMUG 4 EU

igelu International Group of Ex Libris Users

NEWSLETTER FOR SFX/METALIB USERS

Design Ingeborg Kleijnjan for ProBiblio

Photo by: "drookup" Scott Fitzgerald

Photo by: "wartyronin"

Photo by: "fitzgene"

In this issue:

- Integration of Ex Libris products
- Integration in Luxembourg
- Integration with third party products
- Ex Libris and Endeavour merger
- Metalib 4 testing
- e-Books focus group
- MetaLib advanced CKB tools
- Library 2.0

Issue No 4 was edited by

Beate Rusch, KOBV Berlin
Lukas Koster, Library of the University of Amsterdam
Damyanti Patel, Royal Holloway London
Andrea Kroneisl, The Austrian Library Network and Service LTD Vienna
Contact: editors(@)smug-4-eu.org

Hot dogs and cars - The drive-in McLibrary

BY BEATE RUSCH, KOBV BERLIN AND LUKAS KOSTER, LIBRARY OF THE UNIVERSITY OF AMSTERDAM

CONTACT: RUSCH(@)ZIB.DE L.KOSTER(@)UVA.NL

A hot dog vendor expanding into selling hamburgers. Or a car company expanding into tire manufacturing. What do the two have in common? And how does this relate to libraries? Both of them are examples of integration activities – according to Investopedia - horizontal and vertical integration. Horizontal integration means expanding your business into different products that are similar to current lines, while vertical integration is expanding your business into areas that are at different points of the same production path.

If libraries describe their philosophy they'll certainly also talk about integration – integration of services, integration of assets and resources.
Like the car manufacturer, libraries have been expanding vertically from providing content only into providing discovery tools and services, publishing and digitisation.
Like the hot dog vendor, libraries have been expanding horizontally by offering an ever wider range of menus and tastes in tools and services.

With their own vehicles (computers and web browsers) clients are now visiting the McLibrary drive-in to order servings to their own taste from a selection of menus: OPACs, Metasearching tools, OpenURL resolvers, Full text databases, etc.

To make an excellent service all these tools need to function as an integrated system. Not only with each other, but also interacting with the rest, such as the institutional website, the institutional portal, third party's learning environments.

Technically speaking, integration may result in a functioning or a unified whole. This distinction refers to different concepts, which are discussed passionately again and again. A "functioning whole" may be understood as an integrated structure of cooperating tools in which the different parts (like in this case SFX, MetaLib or other tools) can be distinguished as such, whereas a "unified whole" would mean that the end users experience one "new" system or user interface that conceals the integrated parts. In the case of SFX and MetaLib we ask the vendor: how can we integrate SFX and MetaLib with each other and with other Ex Libris tools? For the sake of our patrons we are asking ourselves: what's better: a "functioning whole" (using linking, deep links, common databases, etc.), or a "unified whole" (Primo?, home grown systems using a single user interface and X-server and other API's?)?

Both SFX and MetaLib offer the option of only using the backend via the X-Server API's, which is in fact a form of disintegration. One consequence of this separation of (meta) data and presentation could involve another form of integration, that of combining different databases in real or virtual collections that can be accessed by a number of integrated or non-integrated tools.

When libraries are talking about integration, they are not only thinking about the user's perspective, but also about the backend of their business, where internal workflows are in focus. In general, our aim is to reduce duplicating work, to store the same information only once, to cooperate closely.

By the way: Have you ever thought of the varieties of hot dogs? The different flavours: with salsa, chilli pepper, beans, ketchup, mustard, relish, remoulade. Organic, veggie, kosher. Split, scrambled, smoked. Tastes are endless and (thank God!) no international standard committee is in sight. This is one of the main differences! While we have standards our clients, around the world, have many different tastes. But that might be another story. Maybe the McDrive metaphor is really inappropriate and silly and we libraries are more like a three star restaurant.

Story

BY BEATE RUSCH, KOBV BERLIN AND LUKAS KOSTER, LIBRARY OF THE UNIVERSITY OF AMSTERDAM

CONTACT: RUSCH(@)ZIB.DE L.KOSTER(@)UVA.NL

http://igelu.org/wp-content/uploads/2010/10/smug4eu_issue4.pdf

If *libraries* describe their philosophy they'll certainly also talk about integration – *integration of services, integration of assets and resources*.

To make an excellent service all these tools need to function as an *integrated system*. Not only with each other, but also interacting with the rest, such as the institutional website, the institutional portal, third party's learning environments.

Technically speaking, integration may result in a *functioning* or a *unified whole*.

A "*functioning whole*" may be understood as an *integrated structure of cooperating tools* in which the different parts (like in this case SFX, MetaLib or other tools) can be distinguished as such, whereas a "*unified whole*" would mean that the end users experience *one "new" system* or user interface that conceals the integrated parts.

Combining different databases in *real or virtual collections* that can be accessed by a number of *integrated or nonintegrated tools*.

When libraries are talking about *integration*, they are not only thinking about the user's perspective, but also about the *backend* of their business, where *internal workflows* are in focus. In general, our aim is to *reduce duplicating work*, to *store the same information only once*, to *cooperate closely*.

Libraries – integration of services, assets and resources.

Integrated systems

- *Functioning whole*: an integrated structure of cooperating, distinguishable tools
- *Unified whole*: one new system of concealed integrated parts

distributed, linked

aggregated, silos

Combined databases

- *Real* collections
- *Virtual* collections

aggregated, silos

distributed, linked

Data management

- Harvest data
- Big data
- Distributed harvested indexes

aggregated, silos

aggregated, linked

Backend integration of internal workflows

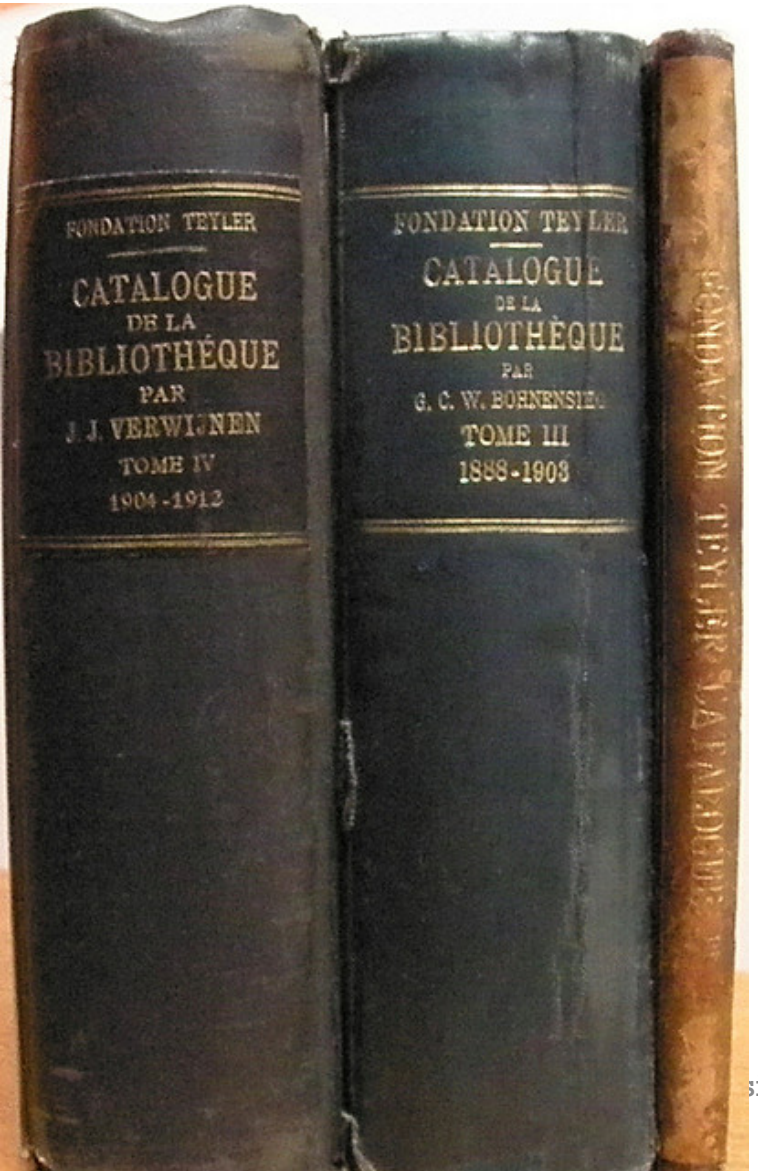
- Reduce duplicating work
- Store the same information only once
- Cooperate closely

aggregated, linked

Story

Libraries: local collections of physical printed publications

http://www.flickr.com/photos/88526197@N00/4218548045/



Catalogues: finding publications and their locations

v.12 Item # 3200000330020 ONLY

Unlinked

DA750 .S5 Ser.2 ~~v.5,~~ 12

Macphail, James Robert Nicolson, 1858–1933, *ed.*

Highland papers ... edited by J. R. N. Macphail, K. C. Edinburgh, Printed by T. and A. Constable for the Scottish history society, 1914~~–34~~. 1916.

Library has v.1–2

2 ~~4~~ v. fronts. (v. 3–4; ports.) fold. facsims. 23cm. (*Half-title:* Publications of the Scottish history society. 2d ser. vol. v, XII, ~~XX; 3d ser. vol. XXII~~)

Vol. 4 contains a biographical introduction by W. K. Dickson.

v.5 done & mark holdings made for both

1. Highlands of Scotland—Hist.—Sources. 2. Highlands of Scotland—Geneal. I. Dickson, William Kirk, 1860– II. Title.

14–11585 Revised 2

Library of Congress DA750.S25

[r35d2]

http://www.flickr.com/photos/cat-sidh/457060770

27 MAY 2012 WIKIPEDIA PUBLIC LIBRARY 12:34pm

Serials

You can search by any of the methods listed here.
Enter the number of the search you want.

1. Serial Title Keyword
2. Serial Title Alphabetical
3. TITLE Alphabetical
4. TITLE Keyword
5. AUTHOR Authority
6. SUBJECT Authority
7. SUBJECT Authority Keyword
8. SERIES Authority
9. LCCN
10. ISBN/ISSN
11. SISAC
12. OCLC #
13. Dewey Decimal CALL #
14. Dynix BIB #
15. Item BARCODE
16. Print Saved Bibliography
17. Quit Searching

Enter your selection(s) and press <Enter>
S=Shortcut on, BB=Bulletin Board, ?=Help

http://en.wikipedia.org/wiki/File:Screenshot_of_Dynix_library_automation_software_in_green.png

http://www.flickr.com/photos/dukeyearlook/7046582583

Library of the University of Amsterdam

Catalogue

Acquisitions | E-shelf | Search History | Results List | My Account | End Session | Help | Ask a question |

Search Advanced Search Browse

Words adjacent?

Words from title — No Yes

AND OR NOT

Author — No Yes

AND OR NOT

Subject — No Yes

Go Clear

Limit search to:

All Locations
ACTA
Allard Pierson Museum
Art History / Cultural studies
Artis Library
Bètawetenschappen
Bungehuis
Bushuis
CEDLA
Emma Kinderziekenhuis
IVIR
UBA-Boekendepot (IWO)
Law library
Medical Library
NEV
P.C. Hoofthuis

(Use *Ctrl* key to select more locations.)

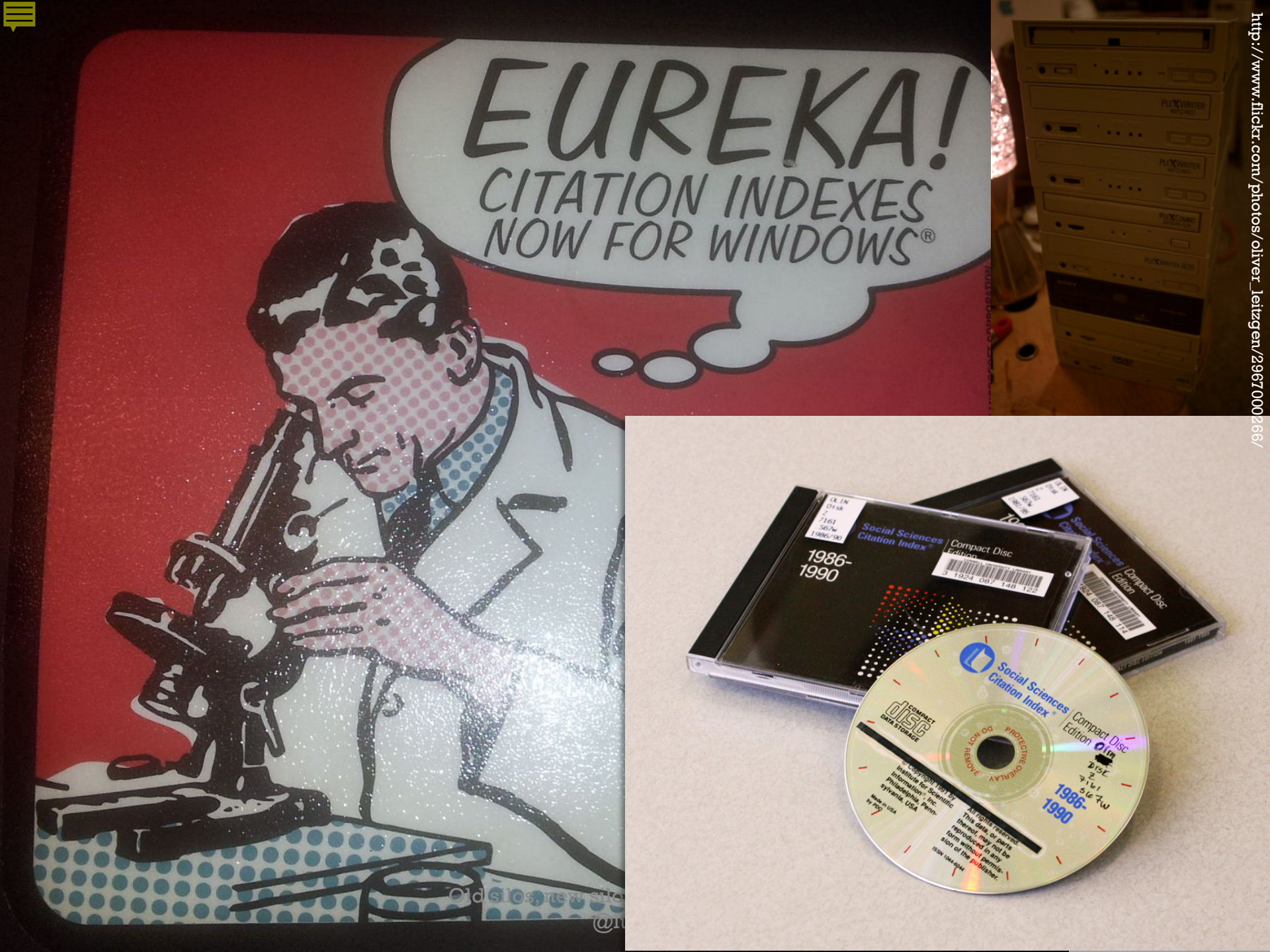
All Languages
Dutch
English
German
French
Spanish
Italian
Russian
--- Scroll down for more ---

(Use *Ctrl* key to select more languages.)

Period: from (yyyy) to (yyyy)

All Formats
Digital
E-books
E-journals
Software
Online resources
Non digital
Books
Manuscripts
Pictures
Assignments and Papers
Audio-visual
Maps
Periodicals / Serials
Sheet music
Sound
Museum specific objects

(Use *Ctrl* key to select more formats.)



http://www.flickr.com/photos/oliver_leitzgen/2967000266/

Netherlands Consortium

MORE INFORMATION FOR NEW USERS

Web of Science®

Search

[ ] in Topic

Example: oil spill* mediterranean

AND [ ] in Author

Example: O'Brian C* OR OBrian C*

Need help finding papers by an author? Use Author Search.

AND [ ] in Publication Name

Example: Cancer* OR Journal of Cancer Research and Clinical Oncology

Add Another Field >>

Search | Clear | Searches must be in English

Current Limits: (To save these permanently, sign in or register.)

Timespan

- All Years (updated 2012-11-02)
- Date Range
  - From: YYYY-MM-DD to: 2012-11-06
  - Use Processing Date instead of Publication Date

Citation Databases

- Science Citation Index Expanded (SCI-EXPANDED) --1975-present
- Social Sciences Citation Index (SSCI) --1975-present
- Arts & Humanities Citation Index (A&HCI) --1975-present

Adjust your search settings

Adjust your results settings

View in: | 简体中文 | English | 日本語

© 2012 Thomson Reuters | Terms of Use | Privacy Policy | *Please give us your feedback on using Web of Knowledge.*

Get EndNote X6 Now!

Store your references and PDFs and find full text in seconds with EndNote X6 and EndNoteSync. Try it now

Training and Support

- Download Quick Recorded Training

Find us on Facebook Like

What's new in Web of Knowledge?

- Data Citation Index℠: Discover, use and cite research data. More Information.
- More of What's New

Featured Tips

- Visualize citation connections at a glance with Citation Mapping (view demo).
- Identify citation trends graphically with Citation Report (view demo).
- How to update your Researcher ID profile.

Customize Your Experience

Sign In | Register

- Save and manage your references online with *EndNote Web* – freely available and fully integrated.
- Save and run searches
- Choose your start page
- Want to know more?
- Sign In to Web of Knowledge to get to ResearcherID.

Customer Feedback and Support

- Contact Customer Support
- Provide Feature Feedback
- Request a Data Change

Silos

Data traps

Silo origins

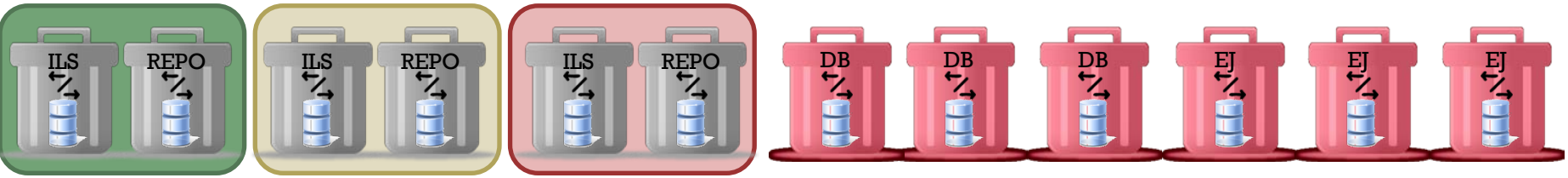
Technology
Economics
Licenses
Tradition
Psychology

http://www.flickr.com/photos/kotomi-jewelry/5035479051/

Serious Story

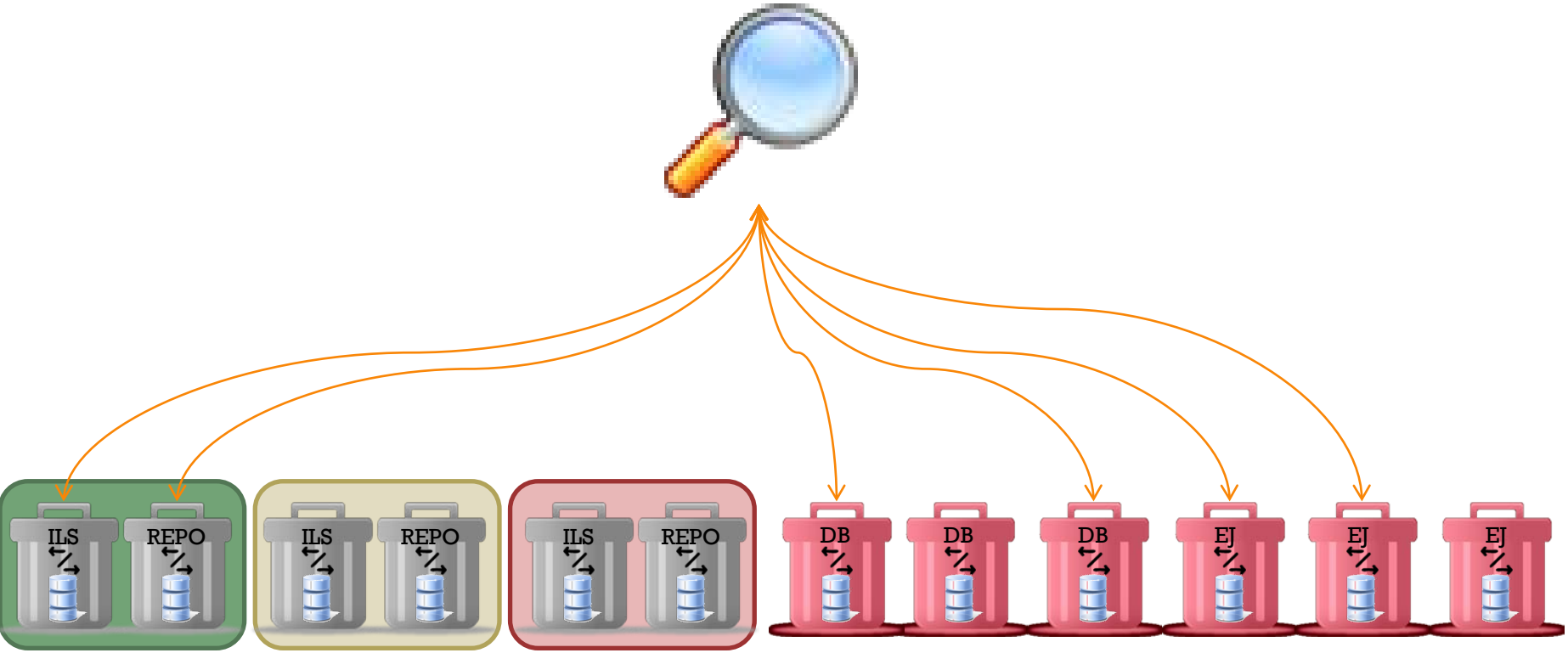
ILS = Library Catalogue
REPO = Institutional Repository
DB = Database
EJ = EJournal

Bibliographic data



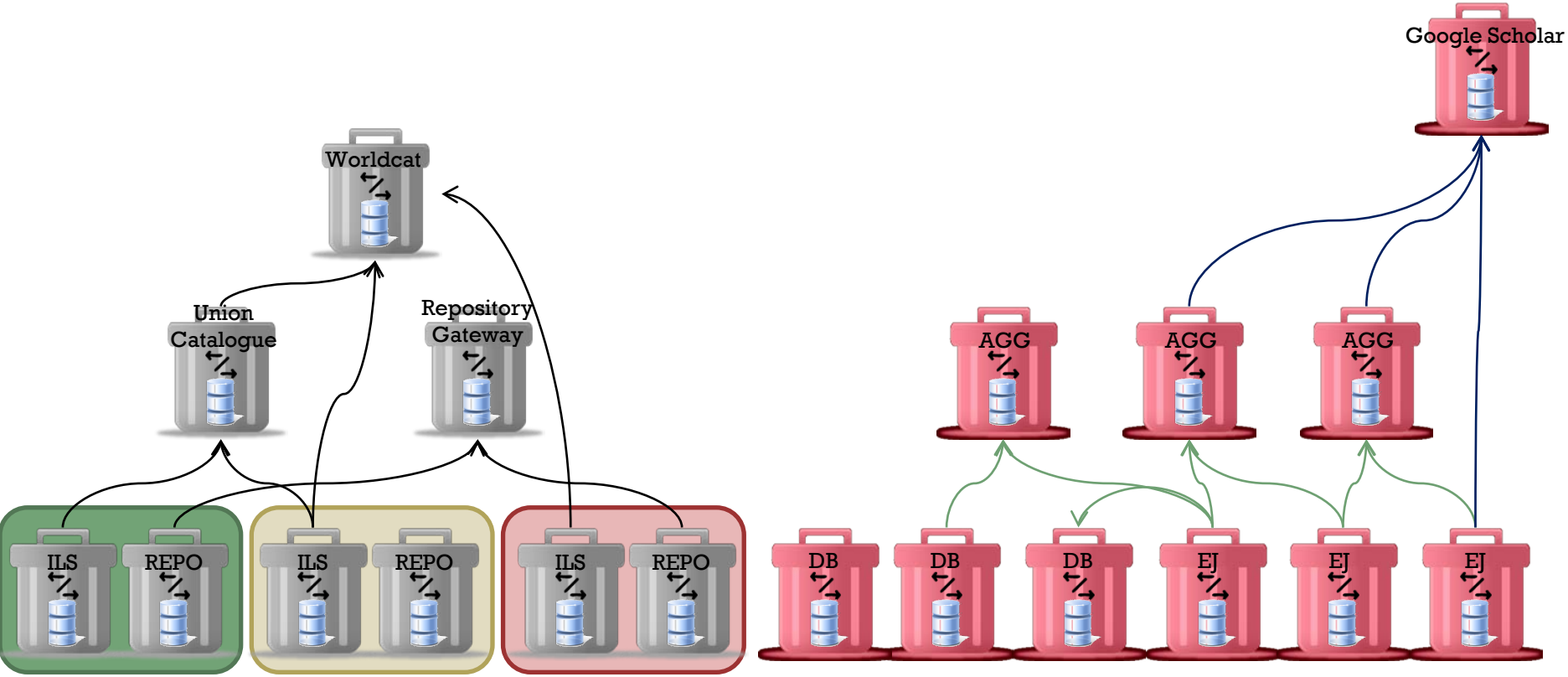
ILS = Library Catalogue
REPO = Institutional Repository
DB = Database
EJ = EJournal

Federated search



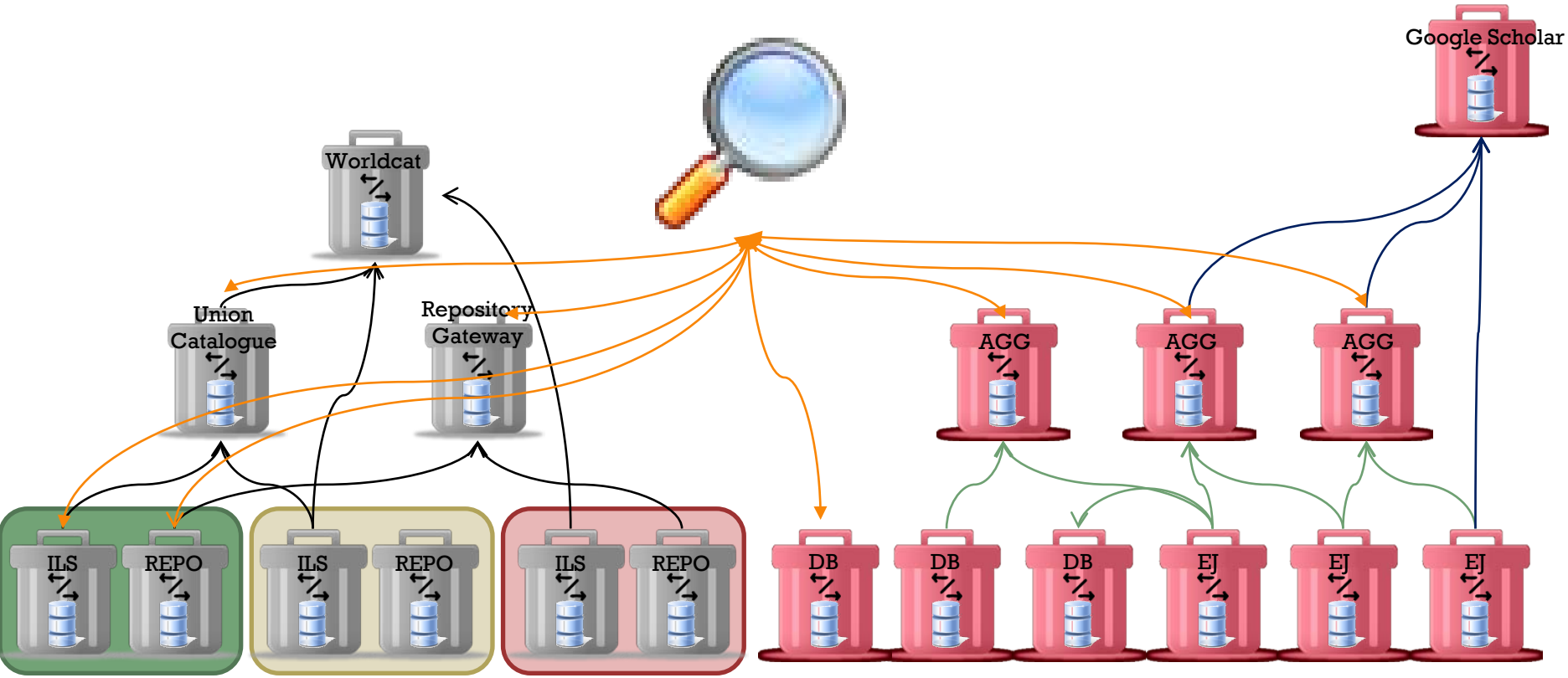
ILS = Library Catalogue
REPO = Institutional Repository
DB = Database
EJ = EJournal
AGG = Aggregation

Aggregations



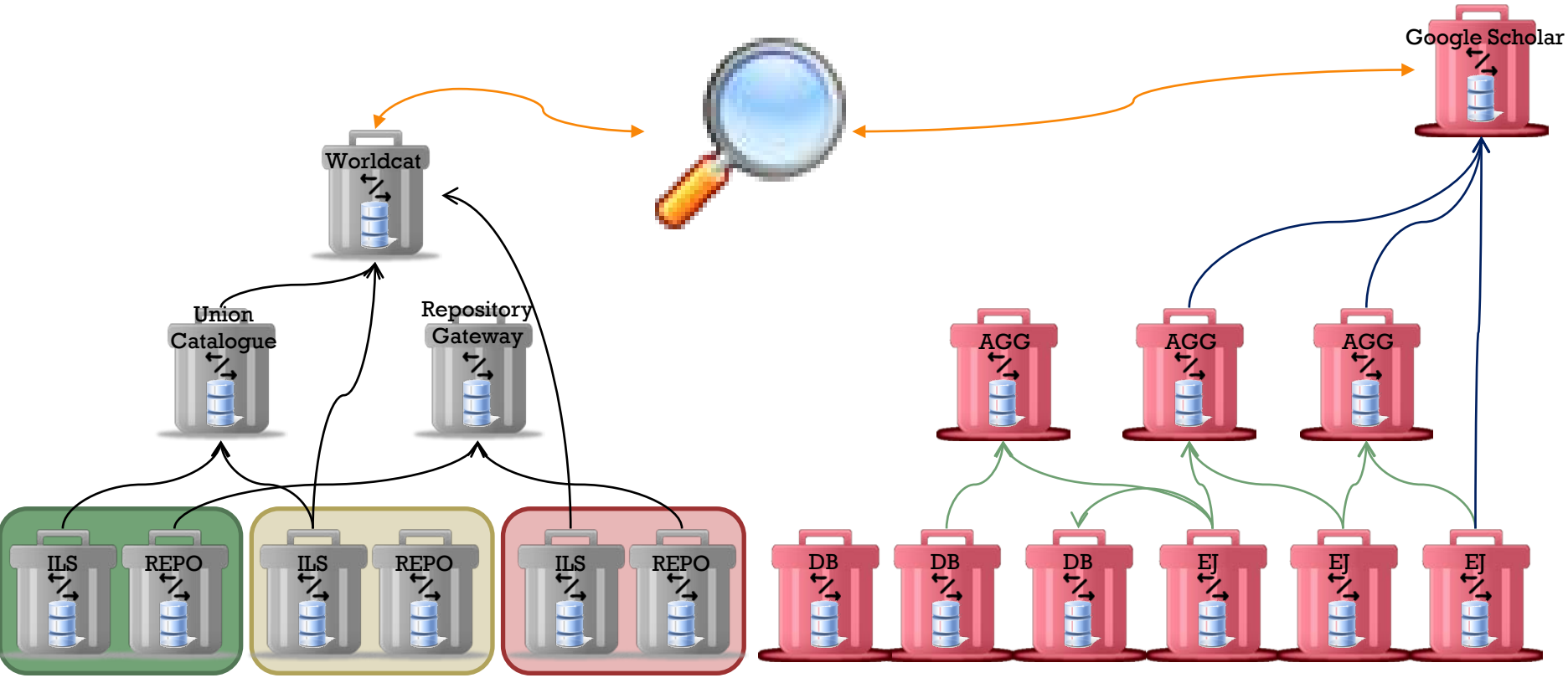
ILS = Library Catalogue
REPO = Institutional Repository
DB = Database
EJ = EJournal
AGG = Aggregation

Federated search

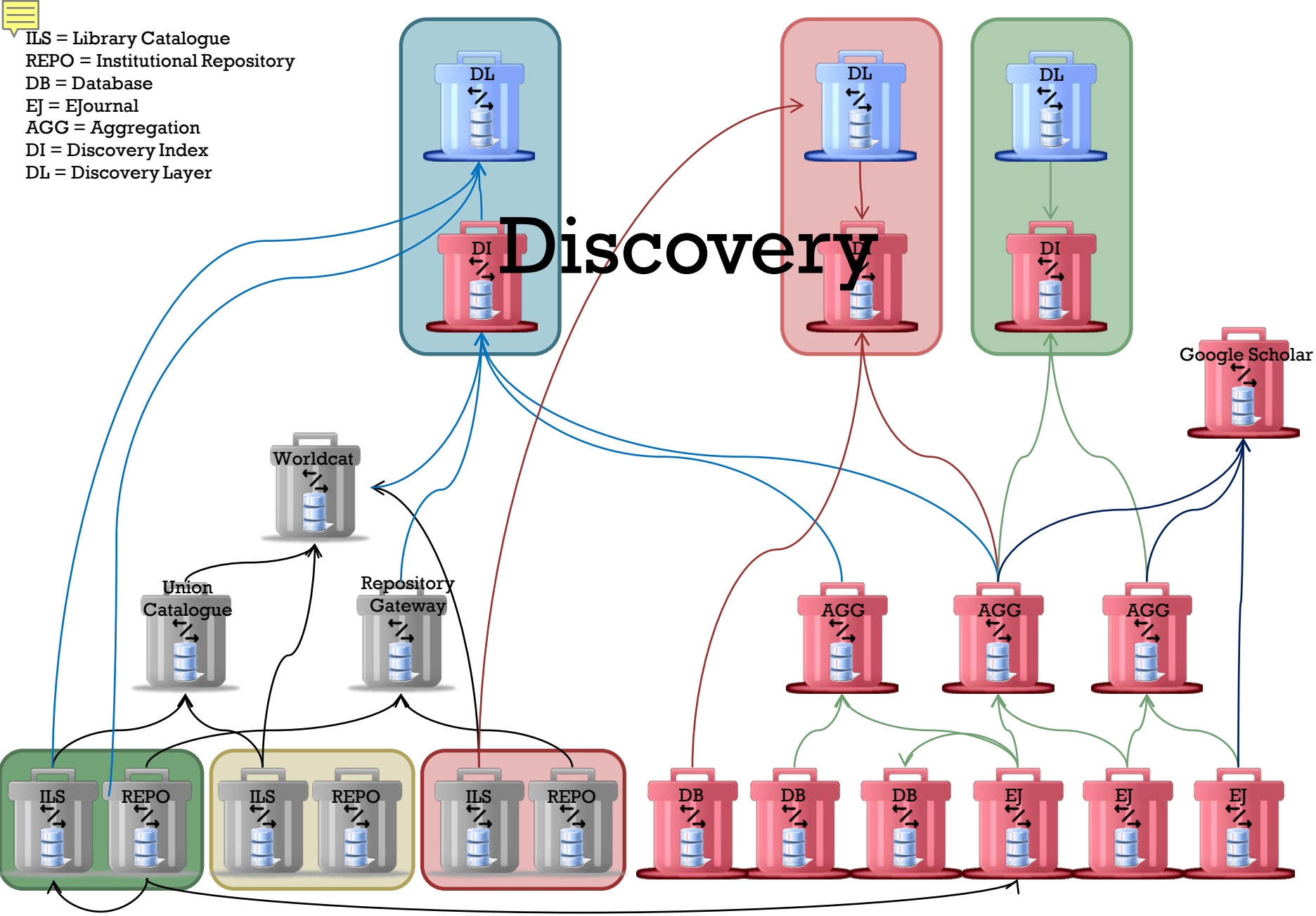


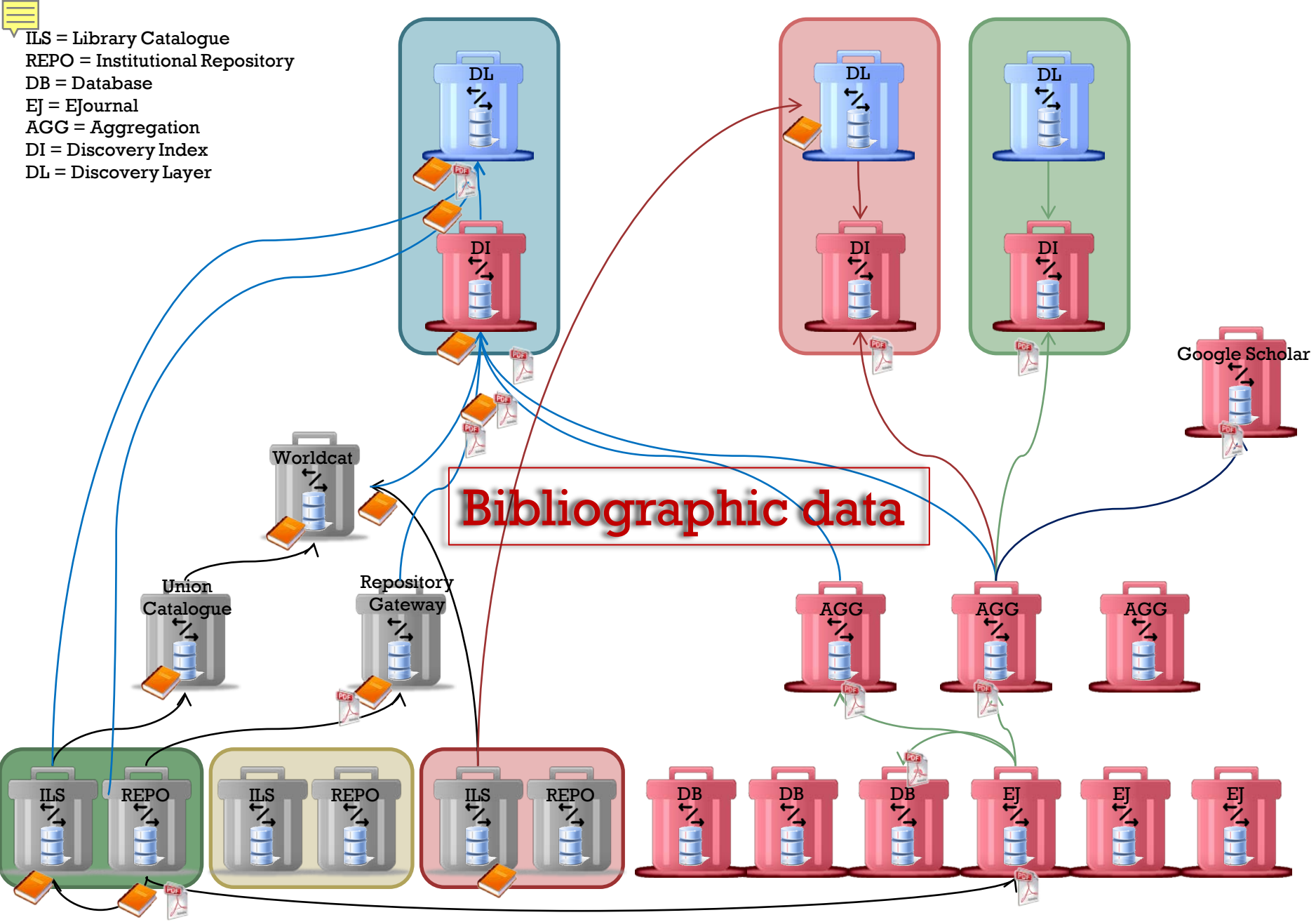
ILS = Library Catalogue
REPO = Institutional Repository
DB = Database
EJ = EJournal
AGG = Aggregation

Aggregated search

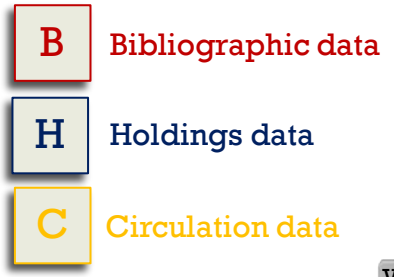


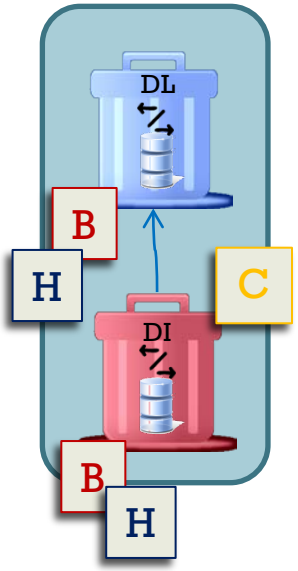
Discovery

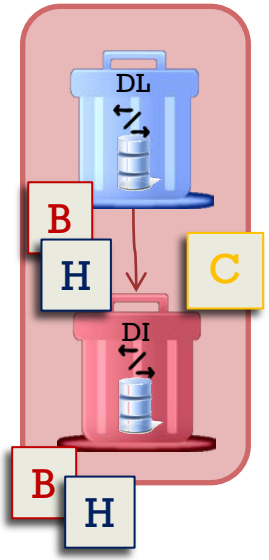

ILS = Library Catalogue
REPO = Institutional Repository
DB = Database
EJ = EJournal
AGG = Aggregation
DI = Discovery Index
DL = Discovery Layer
DL
DI
DL
DI
DL
DI
Google Scholar
Worldcat
Union Catalogue
Repository Gateway
AGG
AGG
AGG
ILS
REPO
ILS
REPO
ILS
REPO
DB
DB
DB
EJ
EJ
EJ

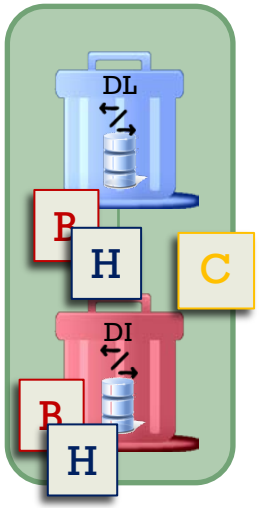

ILS = Library Catalogue
REPO = Institutional Repository
DB = Database
EJ = EJournal
AGG = Aggregation
DI = Discovery Index
DL = Discovery Layer
DL
DI
DL
DI
DL
DI
Google Scholar
Worldcat
Bibliographic data
Union Catalogue
Repository Gateway
AGG
AGG
AGG
ILS
REPO
ILS
REPO
ILS
REPO
DB
DB
DB
EJ
EJ
EJ

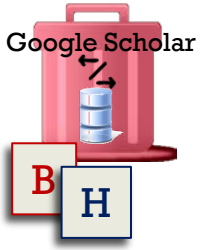
ILS = Library Catalogue
REPO = Institutional Repository
DB = Database
EJ = EJournal
AGG = Aggregation
DI = Discovery Index
DL = Discovery Layer

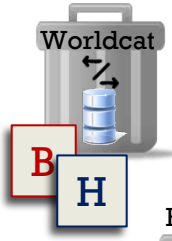


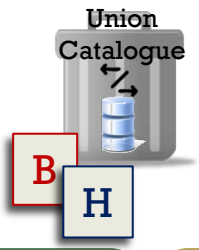


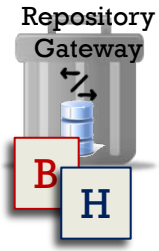


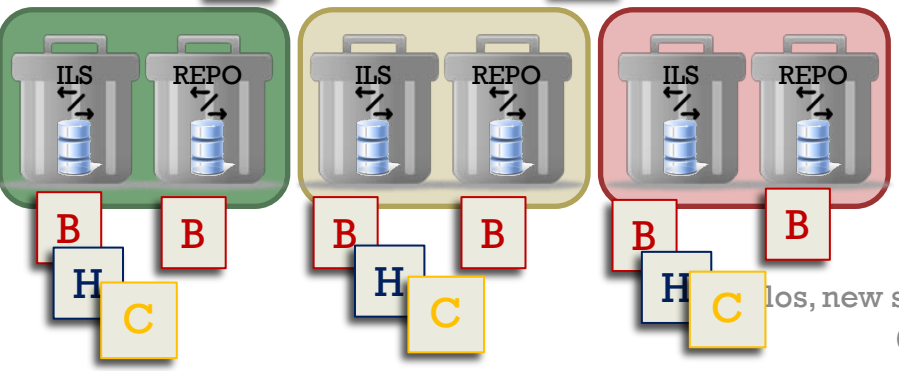


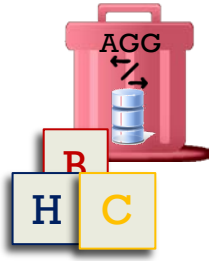


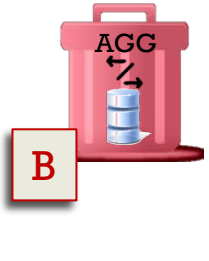


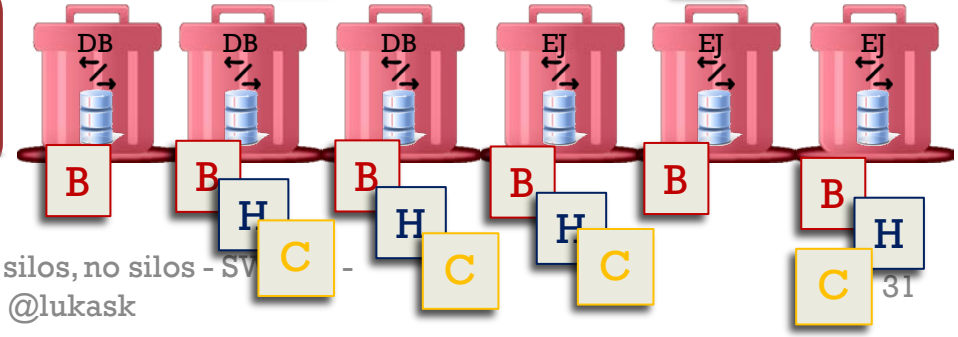










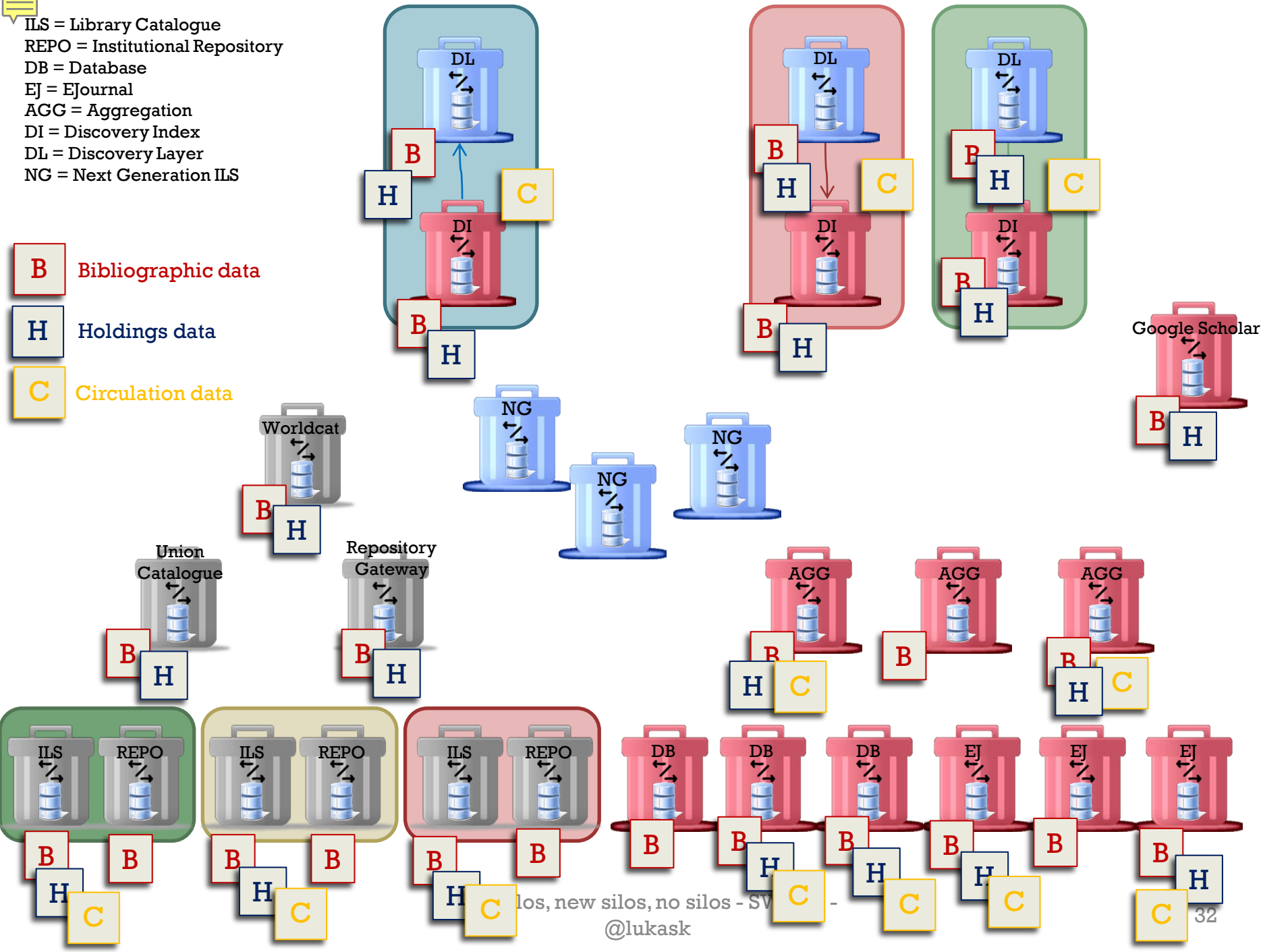


ILS = Library Catalogue
REPO = Institutional Repository
DB = Database
EJ = EJournal
AGG = Aggregation
DI = Discovery Index
DL = Discovery Layer
NG = Next Generation ILS

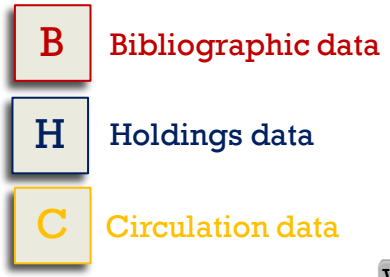
B Bibliographic data

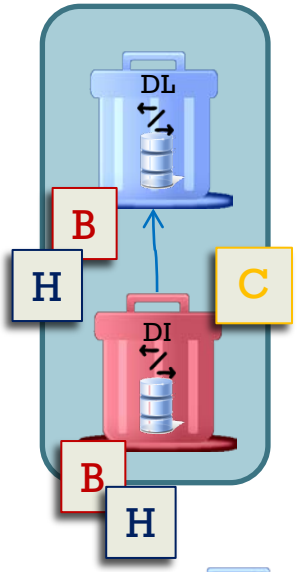
H Holdings data

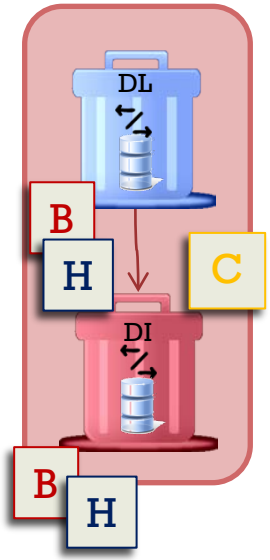
C Circulation data

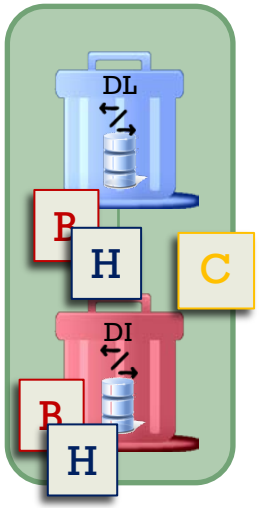


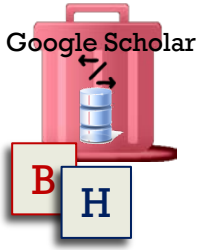
ILS = Library Catalogue
REPO = Institutional Repository
DB = Database
EJ = EJournal
AGG = Aggregation
DI = Discovery Index
DL = Discovery Layer
NG = Next Generation ILS

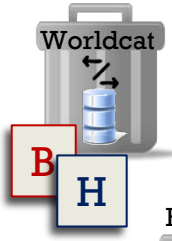


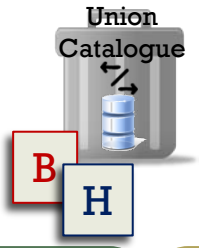


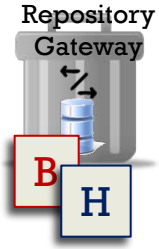


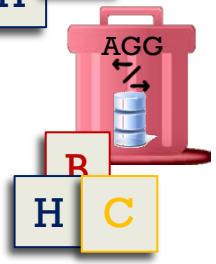


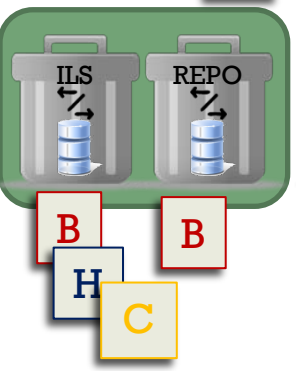


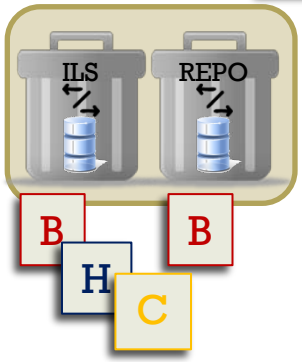


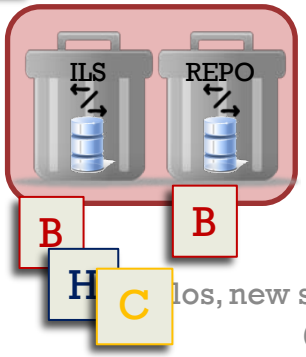


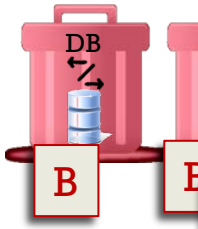


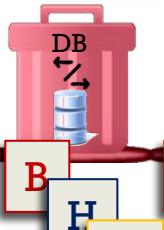


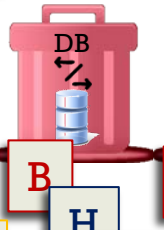


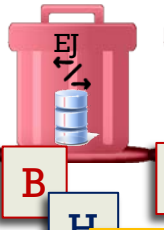


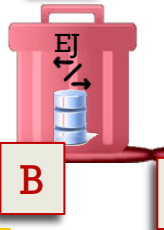










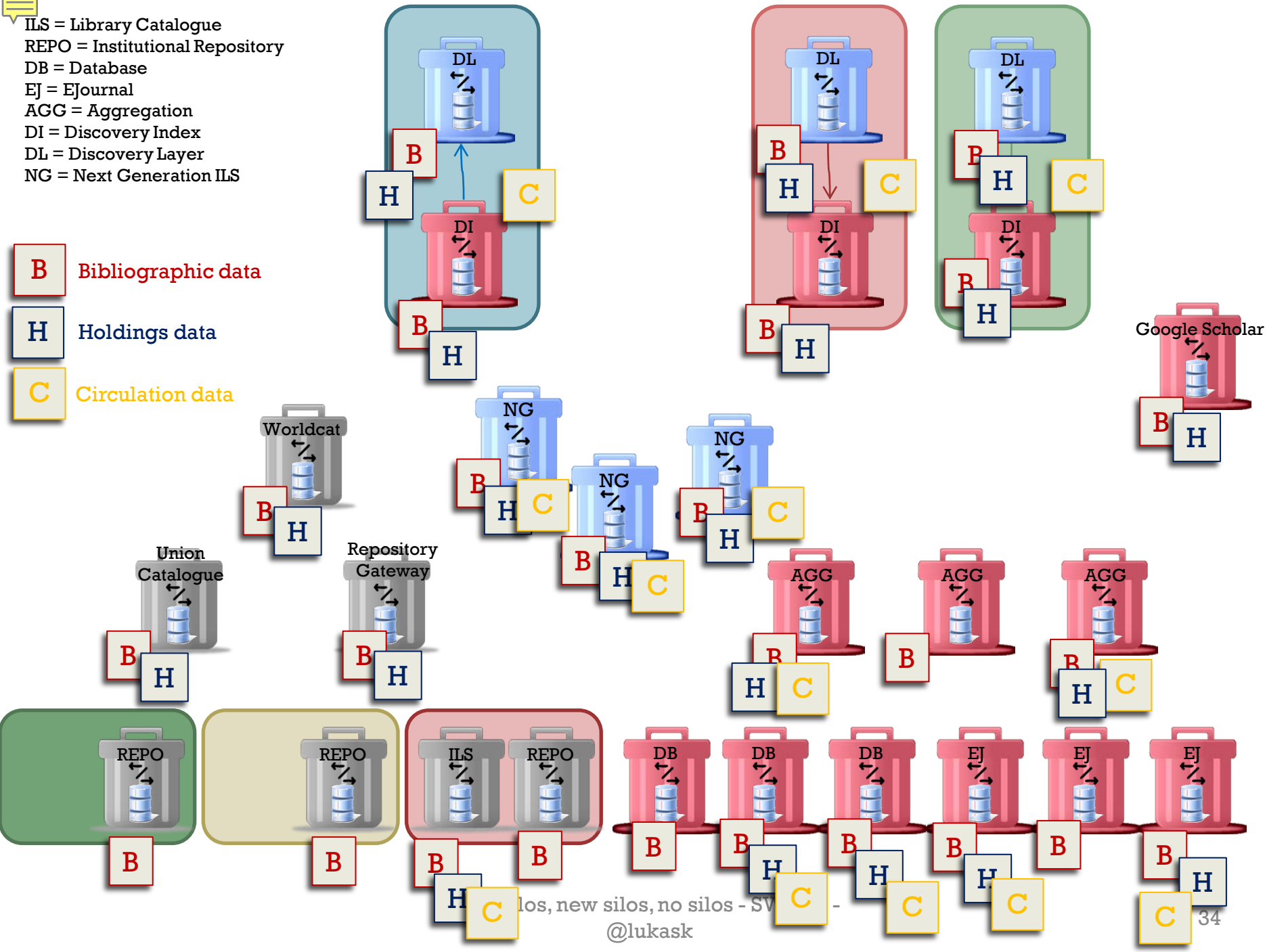


ILS = Library Catalogue
REPO = Institutional Repository
DB = Database
EJ = EJournal
AGG = Aggregation
DI = Discovery Index
DL = Discovery Layer
NG = Next Generation ILS

B Bibliographic data

H Holdings data

C Circulation data



Not very efficient

Free market leads to efficiency?
Not really!

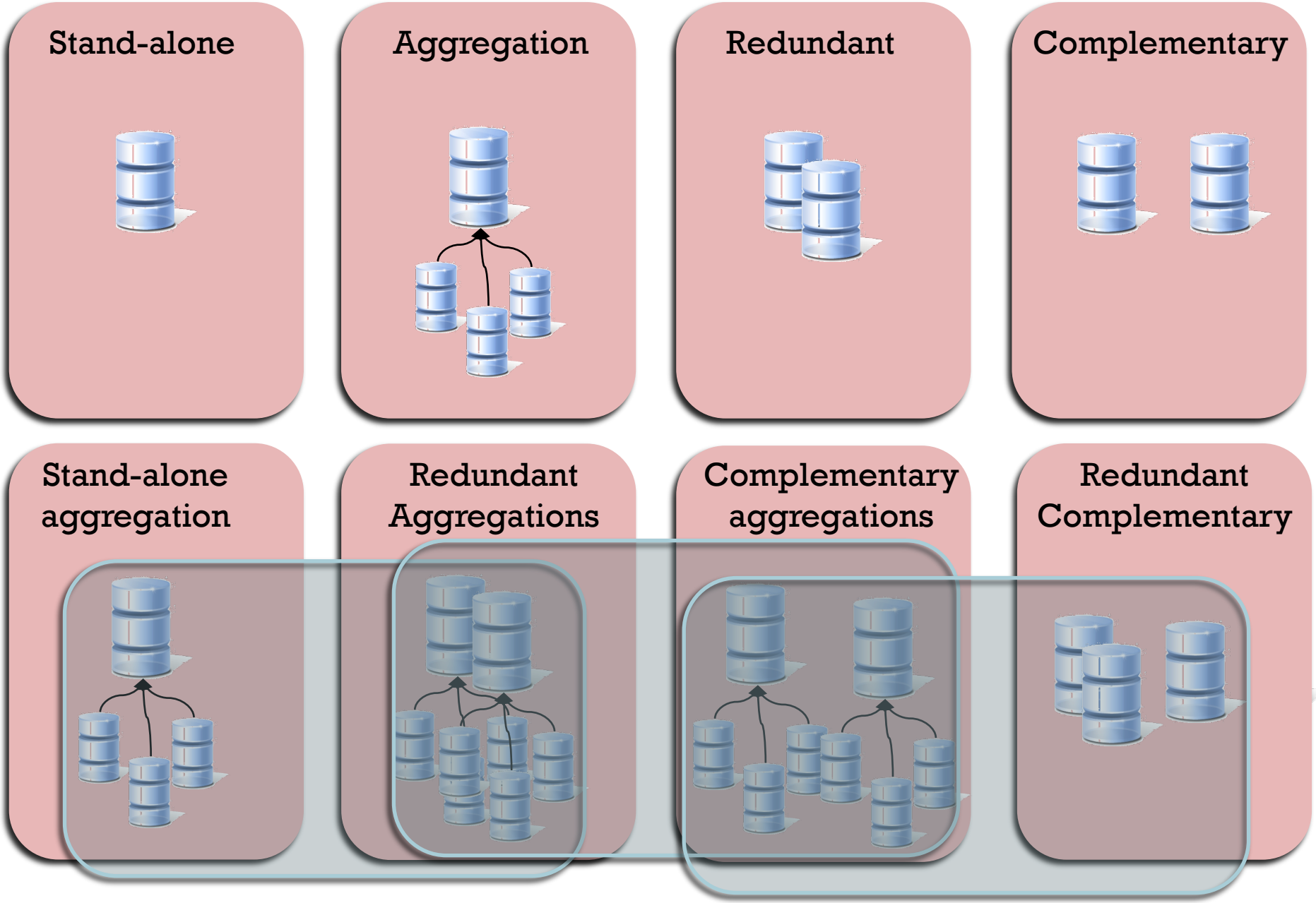
Huge redundancies
Missing data
Obscure, cloudy situation
Unequal access

Library crisis

245 $ a

Silutions?

http://www.flickr.com/photos/thelightningman/5494666930


Stand-alone
Aggregation
Redundant
Complementary
Stand-alone aggregation
Redundant Aggregations
Complementary aggregations
Redundant Complementary

Completely distributed

- Less redundancy
- Up to date
- Trust
- Share workload
- …

- Redundancy risk
- Performance risk
- Persistence risk
- Inconsistent indexes
- Trust
- …

http://en.wikipedia.org/wiki/Fallacies_of_Distributed_Computing

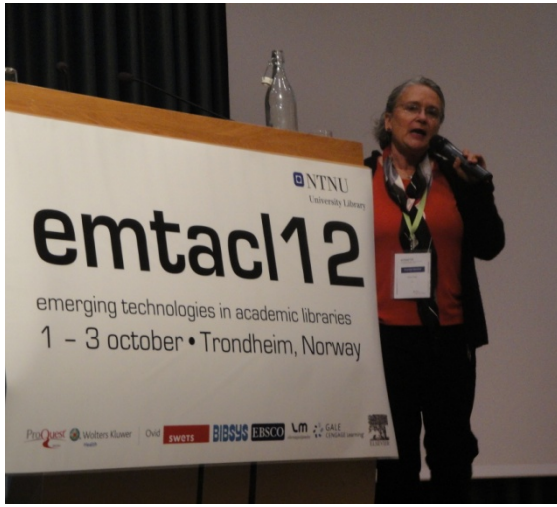
Is linked data the answer?

Karen Coyle at EMTACL12:

"The web is awash in bibliographic data"

"Most of what we would contribute would be duplication of data that already exists"

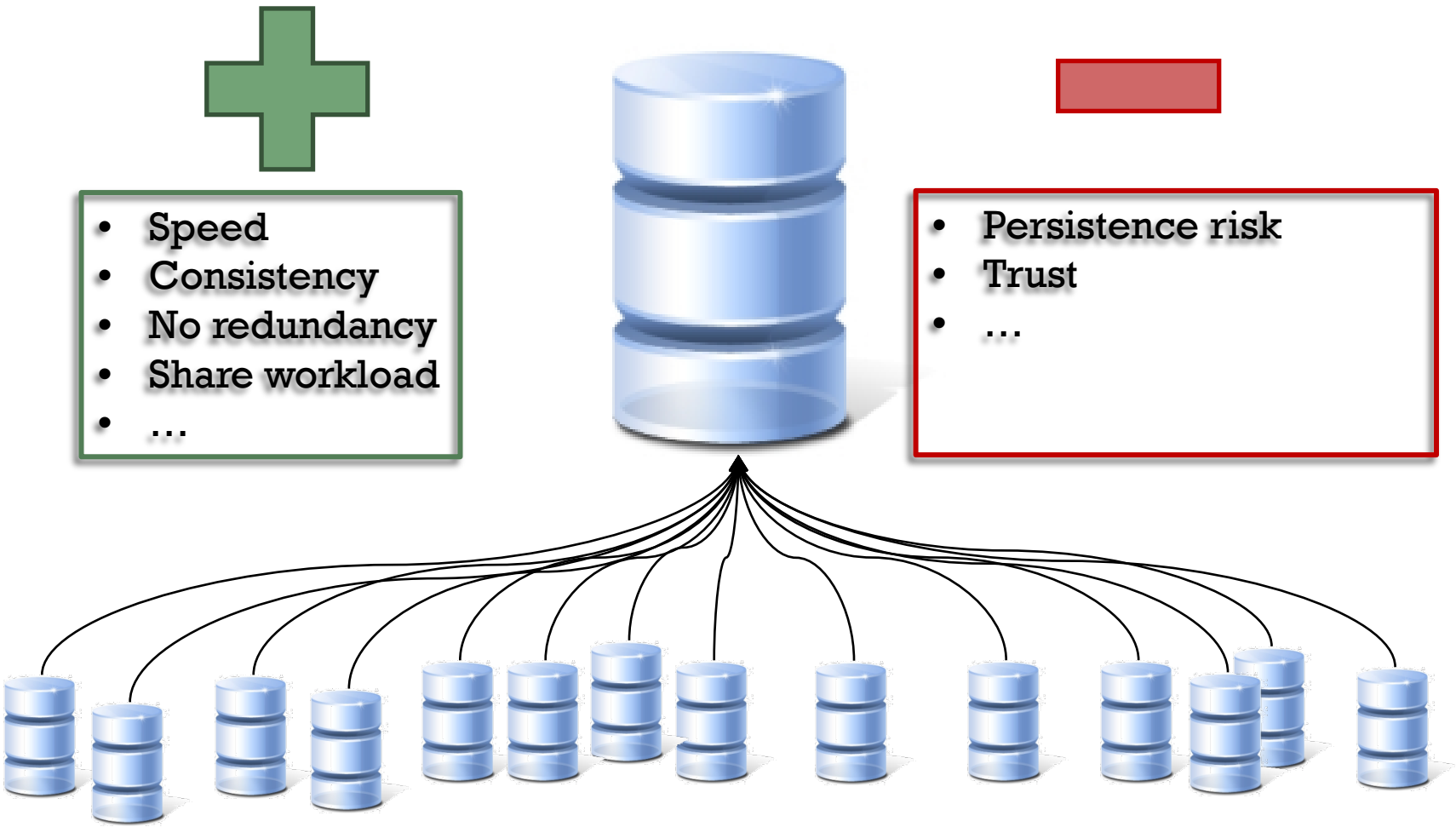
"It is library holdings that is key to providing service and furthering knowledge creation"



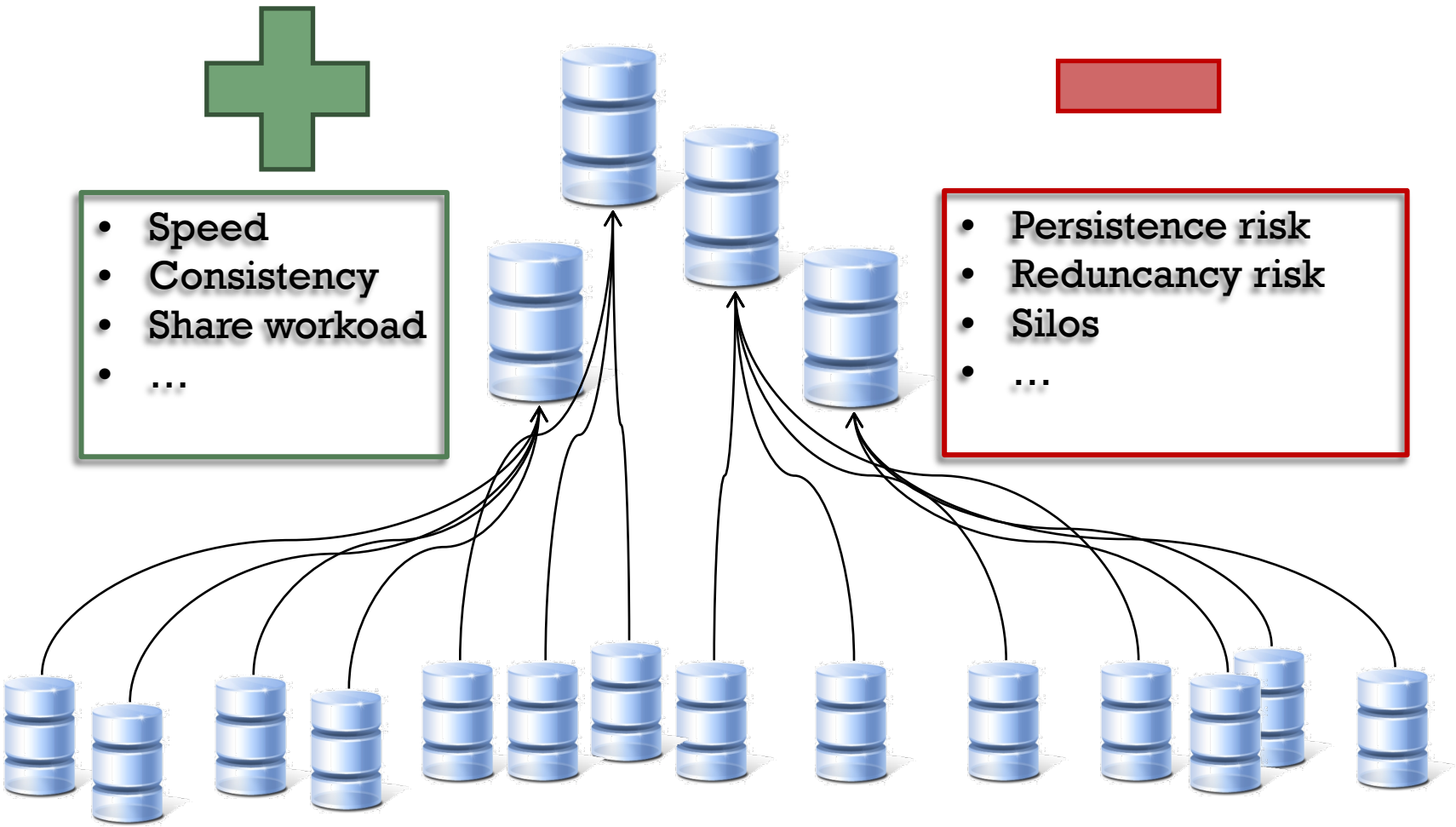
http://www.kcoyle.net/presentations/thinkDiff.pdf

Completely centralised

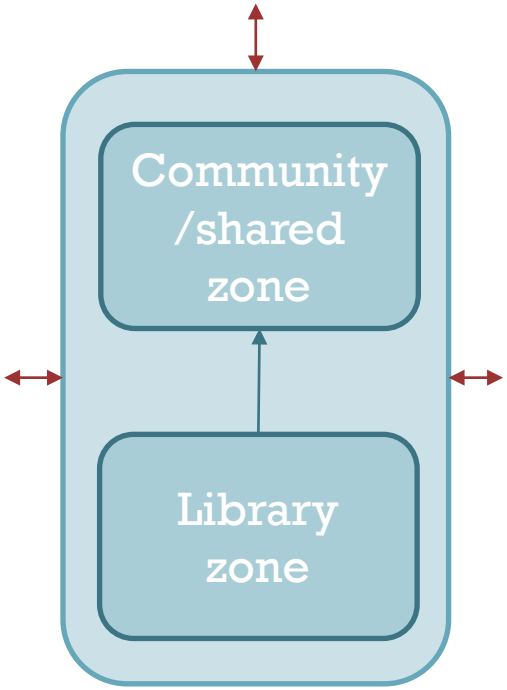
Completely aggregated



Distributed aggregation



Next Generation ILS in the cloud

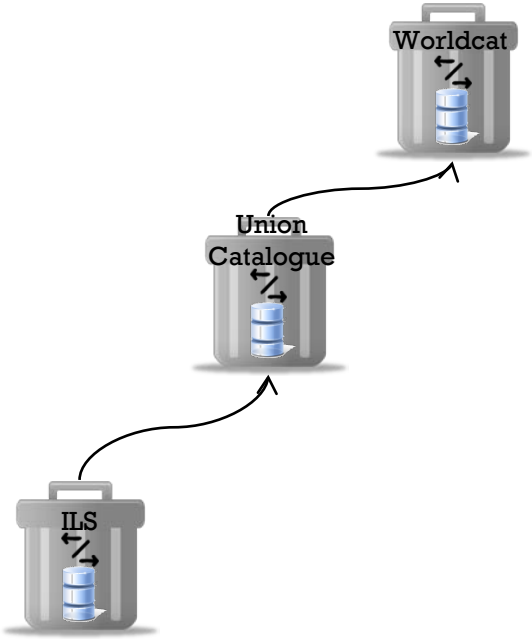


FRBR

Work
Expression
Manifestation

Items

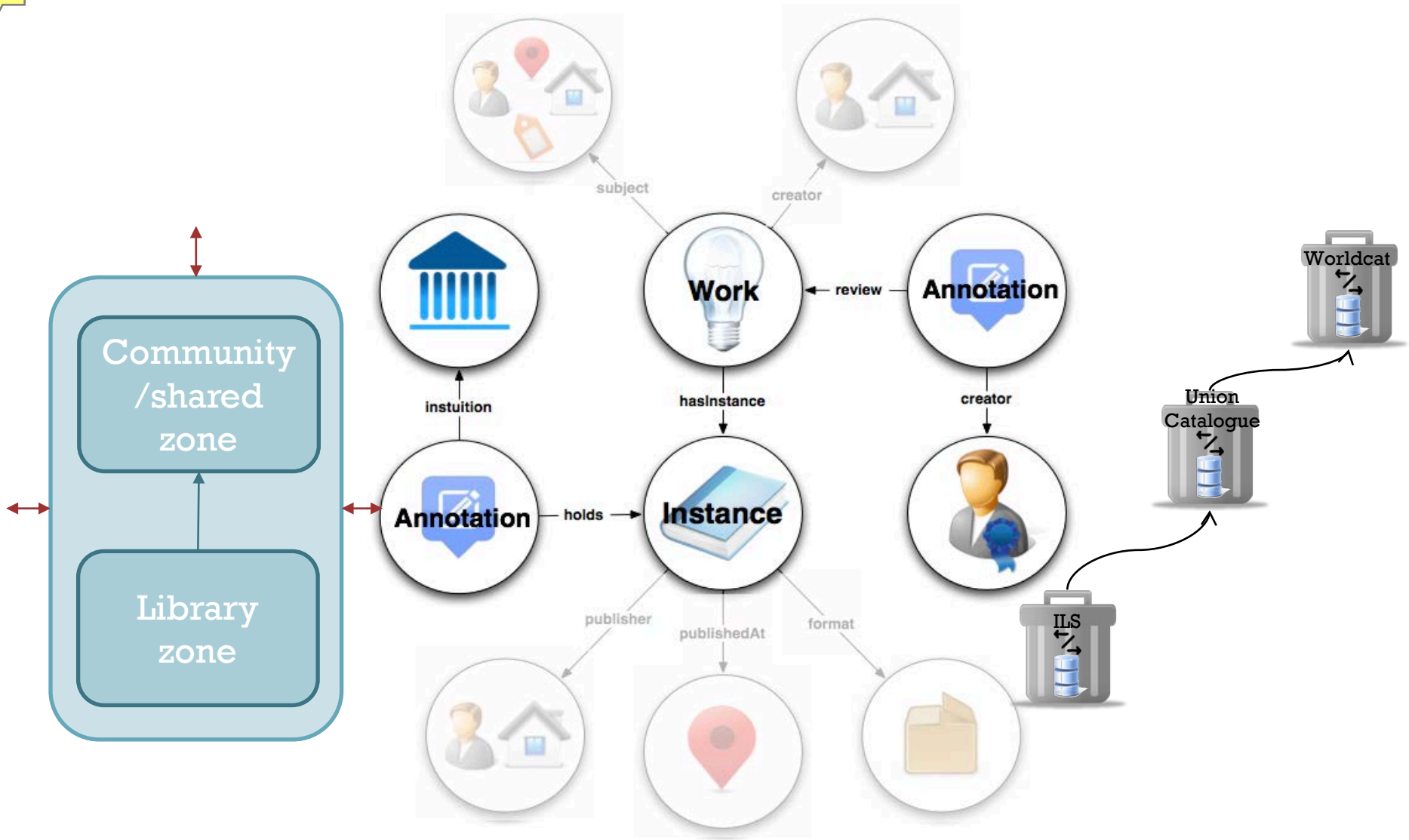
Holdings

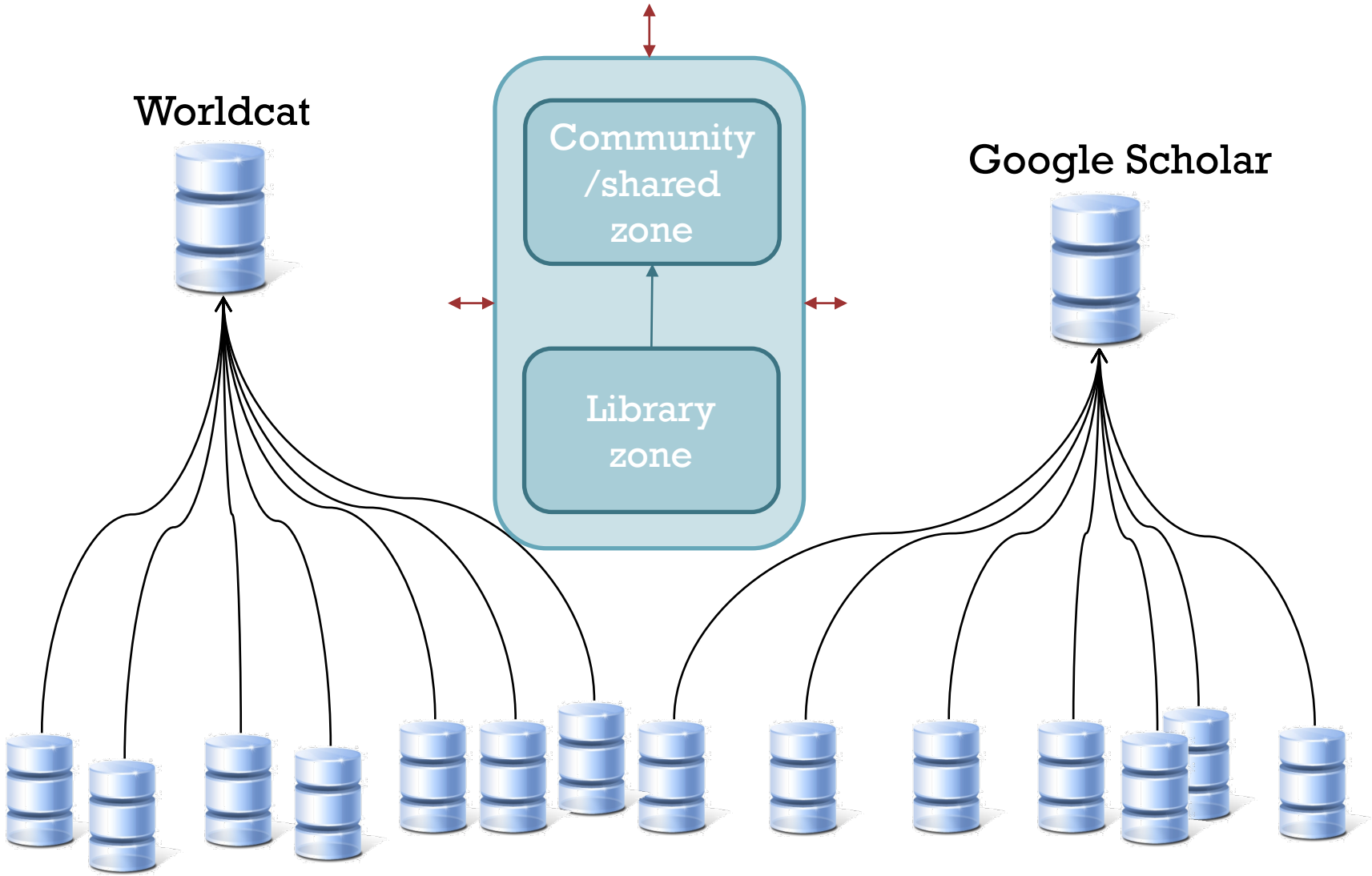


http://thoughts.care-affiliates.com/2012/10/impressions-of-new-library-service.html

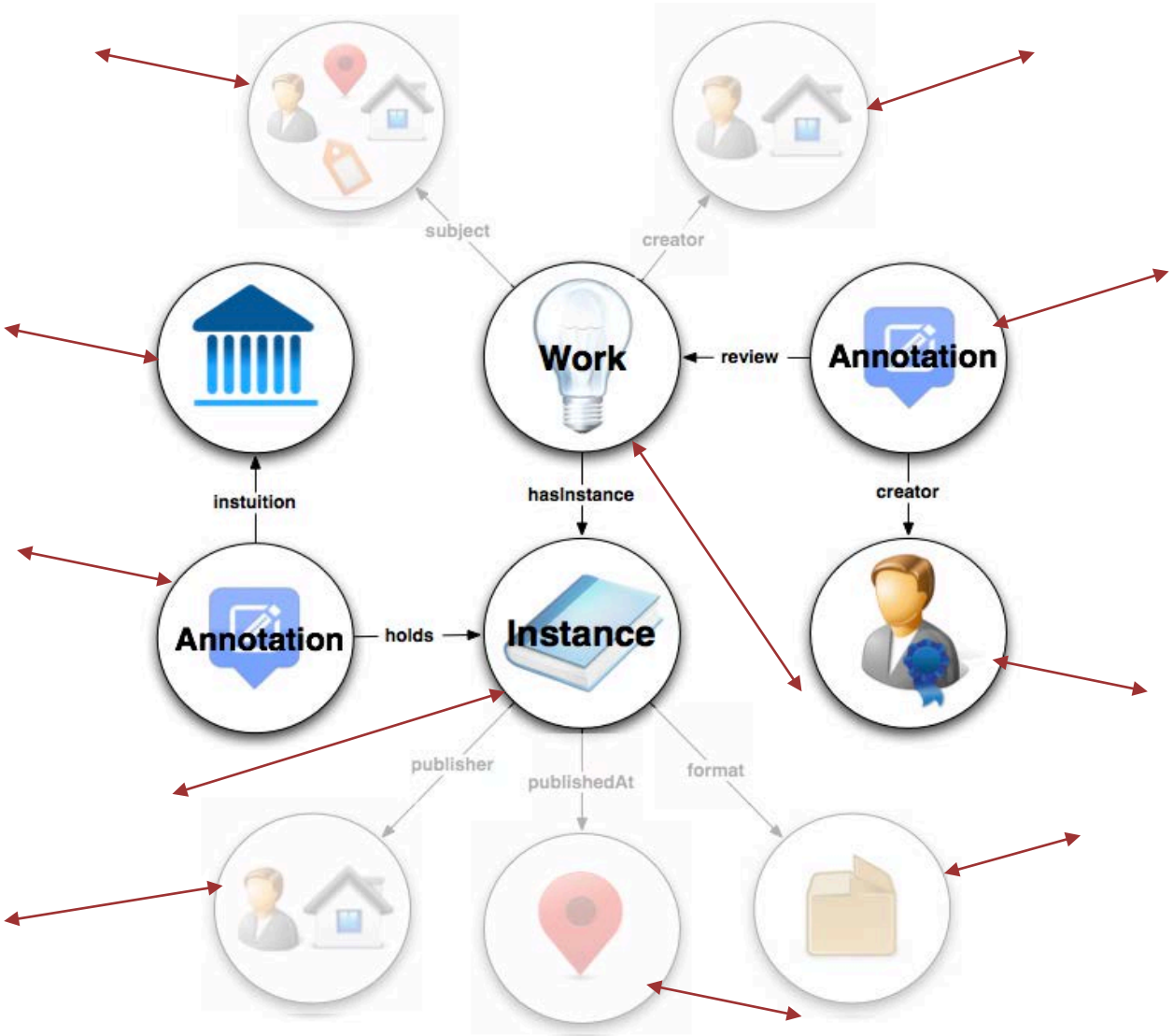
LoC Bibliographic Framework

http://www.loc.gov/marc/transition/


Community /shared zone
Library zone
subject
creator
Work
review
Annotation
instuition
hasInstance
creator
Annotation
holds
Instance
publisher
publishedAt
format
Worldcat
Union Catalogue
ILS


Worldcat
Community /shared zone
Library zone
Google Scholar

Brave new world



Library roles

- Reconsider objectives
- Shift focus
- Communicate with system vendors
- Communicate with content vendors
- Engage in framework transition
- Cooperate

Don’t think systems
Think data