

开放与不开放资源 - 利与弊

“数字资源开放获取就研讨会”

CALIS第十三届引进数据库培训周

谢国伟 **Vincent Cheah**
Senior Product Manager, Gale




Gold Mining 金矿



What are needed for Gold Mining?

获得金矿需要什么？

- A knowledge of where gold has been found in the past
过去的黄金在哪里发现
- A knowledge of where one can legally prospect
哪些地方可以合法开采
- A knowledge of prospecting laws and regulations
法律法规
- A knowledge of gold deposits and geology
金矿和地质
- A knowledge of prospecting methods
开采方法
- A knowledge of right equipment
合适的设备
- A knowledge of right technologies
正确的技术
- A knowledge of skillful gold miners
熟练的黄金矿工
- ... many others



¥ \$

AN
EMBASSY
FROM THE
East-India Company
OF THE
UNITED PROVINCES,
TO THE
Grand Tartar Cham
EMPEROR OF
CHINA,
Deliver'd by Their Excellencies
PETER & GOYER, and JACOB & KEYSER,
At His Imperial City of
PEKING.
WHEREIN
The Cities, Towns, Villages, Ports, Rivers, &c.
in the Provinces
CANTON to PEKING.
Are ingeniously Describ'd,
By **M^r JOHN KIRBUHOFF,** Steward to the
AMBASSADORS.
ALSO
An Epistle of Father **JOHN ADAMS** their Antagonist,
Concerning the Whole Negotiation.
With an **APPENDIX** of several REMARKS taken out of
Father **ATHANASIUS KIRSCHER,**
English'd, and fr. French with these General Supplements,
By **JOHN OGILBY** Esq;
His MAJESTIES Cosmographer, Geographical Printer, and Master of the Royal Society in the
KINGDOM of IRELAND.
The Second Edition.
LONDON,
Printed by the Author at his House in White-Friars. M. DC. LXXIIII.

PROSPECTUS DIRECTORY ADVERTISED.

FIRST-CLASS FIRE INSURANCE.

Market Fire Insurance Co.

37 Wall Street, New York.

Condition of the Company, May 1, 1860.

Gross Assets - \$620,782 67
Capital -
Clear Surplus (Over 100)

In presenting the *Ann* Officers have pleasure in and congratulating the 2 2 months generally, upon its of the year just terminated

Extracts from the *Ann* the Insurance Department
The *Surplus* issued by this people, can be so used for a camp in the following manner
the said *Surplus* Fund, shall amount Five Hundred to equal the average annual towards its redem

Surplus Fund, standing on that condition, we do im—in trust as the Capital.

This and French, several of the Great National Ban do and other cities, that our Policies on the "forti ed twenty per cent, more—than the Policies a called," as it were, by the large *Surplus* fund, has business and prosperity of the Company—be be business and prosperity of the Company—be

The Portugues Asia :

OR, THE
HISTORY
OF THE
Discovery and Conquest
OF
INDIA
BY THE
PORTUGUES;
CONTAINING
All their Discoveries from the Coast of
Africa, to the farthest Parts of *China* and
Japan; all their Battels by Sea and Land,
Sieges and other Memorable Actions; a
Description of those Countries, and many
Particulars of the Religion, Government
and Customs of the Natives, &c.

In Three Tomes.

Written in Spanish by *Manuel de Faria y Sousa* of
the Order of Christ.

Translated into English by *Cap. John Stevens.*


Tomus the First.

LONDON, Printed for *C. Brome*, at the Sign of
the *Gun*, at the *West-End* of *St. Pauls*. 1695.

What are needed for Digital Primary Sources?

获得数字原始资源需要什么？

- A knowledge of where primary sources have been found in the past
历史原始资源在哪里获得
- A knowledge of where one can legally prospect
哪些资源可以合法获取
- A knowledge of prospecting laws and regulations
法律与法规
- A knowledge of unique collection and contents
独特的合集和内容
- A knowledge of prospecting methods
寻找方法
- A knowledge of right equipment
合适的设备
- A knowledge of right technologies
正确的技术
- A knowledge of skillful advisors and researchers
经验丰富的咨询委员和研究人员
- ... many others



¥ \$

Why Digital Archive is valuable and expensive? 为何数字档案更为珍贵？

Let do a case study on 让我们以《每日邮报》数字档案进行案例研究

Daily Mail

《每日邮报》数字档案

Outlines 概述

Sourcing the Material 引用资料

Image Capture 捕捉图片

Optical Character Recognition 光学字符识别（OCR）

Metadata 元数据

XML XML 格式

Quality Assurance 质量保证

Application Build 检索举例

Sourcing the material: Which Edition?

引用资料：哪一版？

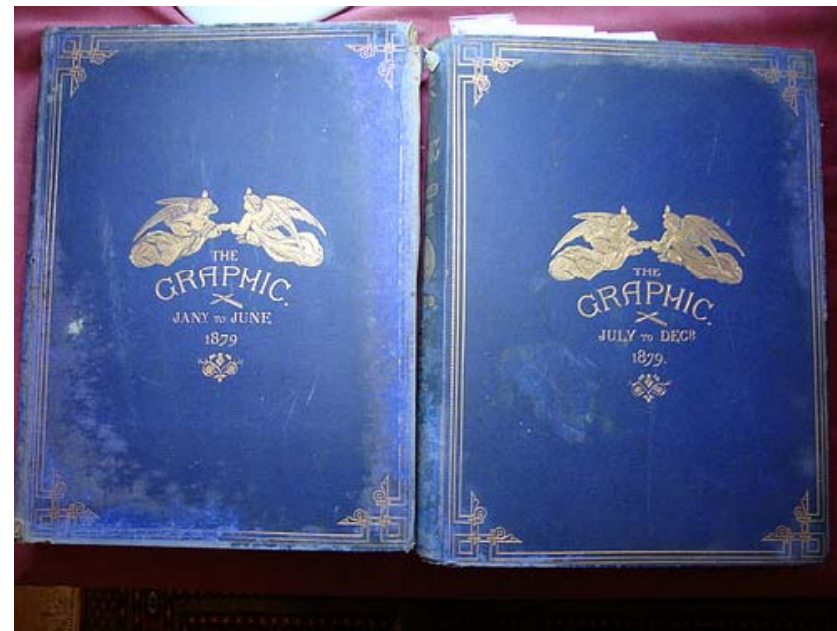


Sourcing the material: Which Edition? 引用资料：哪一版？



- Microfilm?
缩微胶卷？

- Or Hard-Copy Originals?
或原始图书？





MA
DAILY MAIL
ATLANTIC EDN.
A

DAILY MAIL
ATLANTIC EDN.
A

DAILY MAIL
ATLANTIC EDN.
A

DAILY MAIL
ATLANTIC
EDITION
A B O
COMPLETE
1001
TO
1100
STOCK

DAILY MAIL
ATLANTIC EDITION
WIRELESS PAG
N° 1001 TO 1200
STOCK

WIRELESS PAGE
Nos. 901-1000

WIRELESS PAGE
Nos. 901-1000

Nos. 901 TO 1000

WIRELESS PAGE
Nos. 941-1000

STOCK

STOCK

STOCK

STOCK

STOCK

STOCK

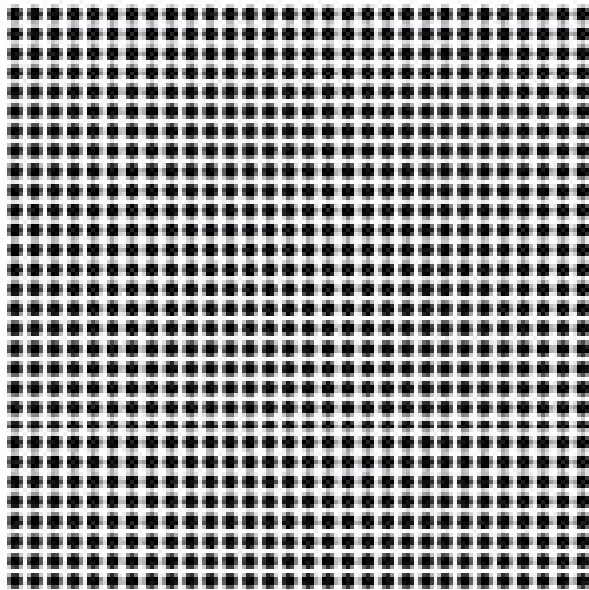
Image Capture 捕捉图片



Image Capture 捕捉图片



72 dots (pixels) in 1 inch
| ← 1 inch → |

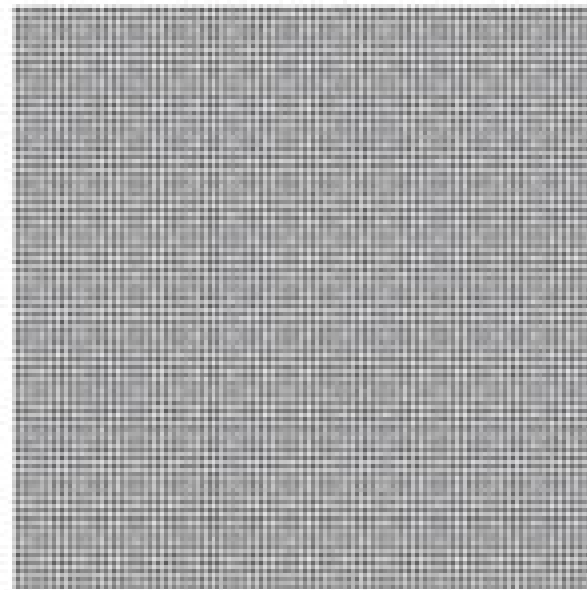


Low Resolution

(72dpi)

低像素

300 dots (pixels) in 1 inch
| ← 1 inch → |



High Resolution

(300dpi)

高像素

Image courtesy of www.zenfolio.com

Bitonal or Greyscale? #1

黑白或灰度? #第一组



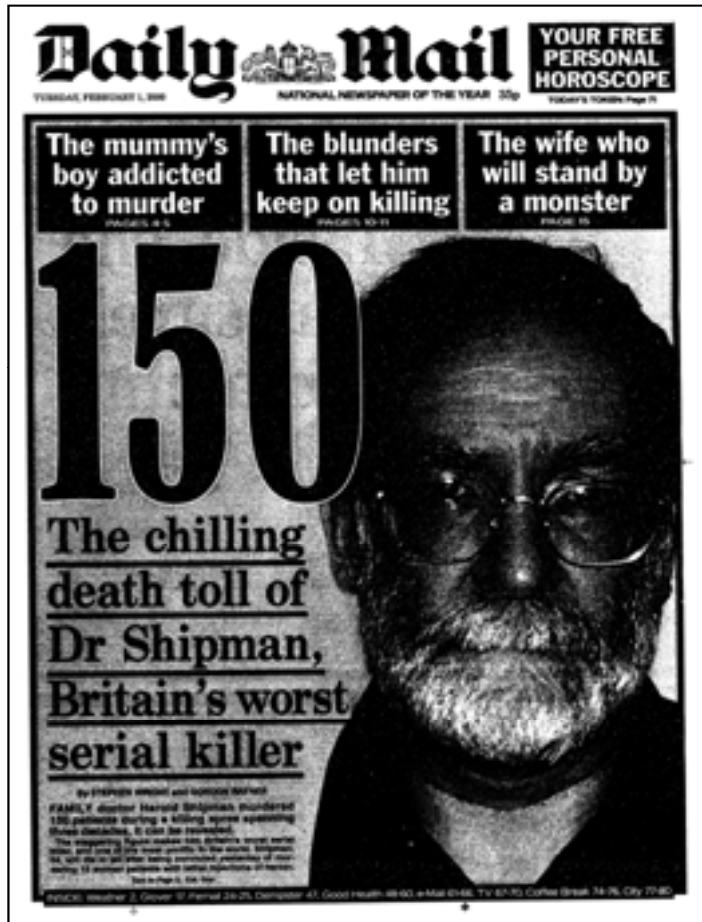
黑白



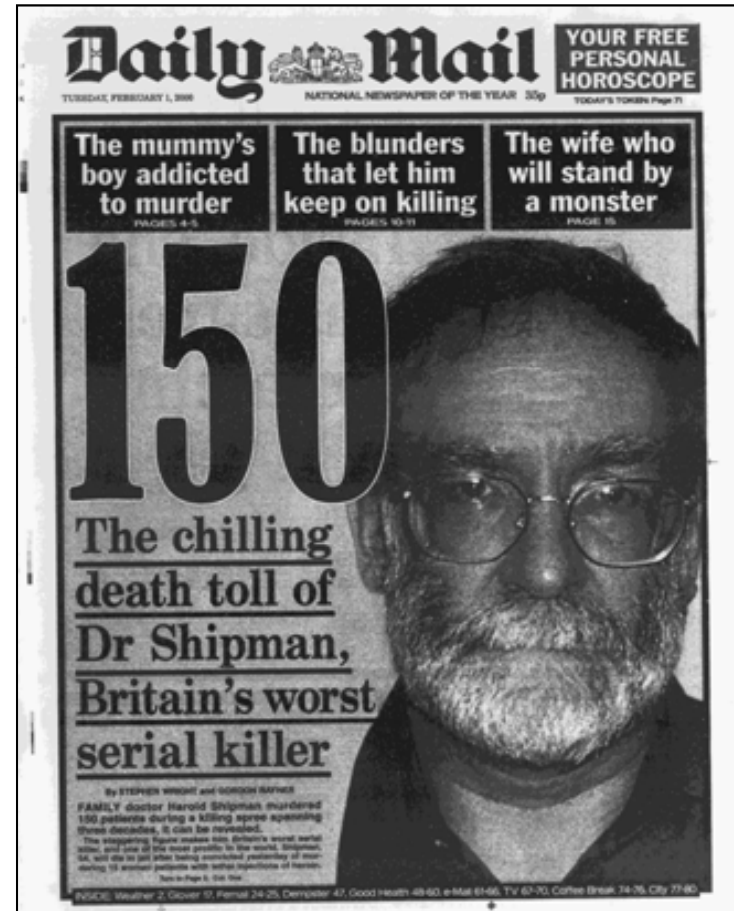
灰度

Bitonal or Greyscale? #2

黑白或灰度？#第二组

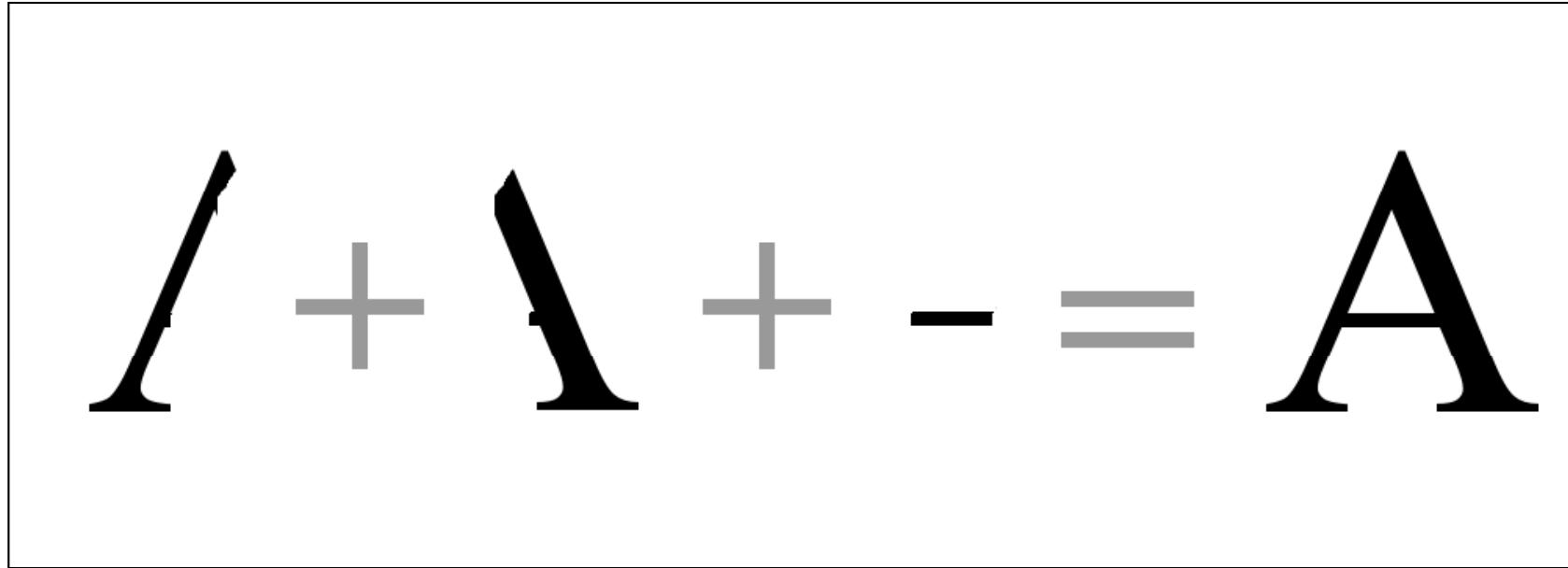


黑白



灰度

Optical Character Recognition (OCR) 光学字符识别 (OCR)



▪ Image courtesy of www.explainthatstuff.com



Who?

- (George Gordon)

人物: George Gordon

What?

- (Article Title: 5AM: Bush Stops War)

- (Article Type: News)

- (679 words)

文章标题: 布什停止战争

文章类型: 新闻

679字



Where?

- (Daily Mail, London, England)

- (Page 1)

- (Issue 29,453)

地点:

《每日邮报》，伦敦

第1页，第29期，453

When?

- (28 February 1991)

Examples of Newspaper Metadata

报纸元数据举例

Article Title 文章标题

Author or Source 作家 或 来源

Newspaper Title 报纸标题

Date of Publication 出版日期

Issue/Edition Number of the Newspaper 期次/版次 报纸编号

Page Number 页码

Article Category (e.g. Advertising, News) 文章分类 (广告、新闻等)

<GALENP>

<Newspaper>

<issue>

<metadatainfo>

<newspaperID>DMHA</newspaperID>

</metadatainfo>

<id>DMHA-1942-0103</id>

<is>14254</is>

<da>January 03, 1942</da>

<pf>19420103</pf>

<dw>Saturday</dw>

<ip>4</ip>

<copyright>© Associated Newspapers Limited</copyright>

<imdim>3784,5357</imdim>

<article type="article">

<id>DMHA-1942-0103-0001-002</id>

<ti>Vast Line-Up against Axis</ti>

<ta>Churchill-Roosevelt 'World Plan' Ready</ta>

<ta>26 Nations Pledge 'No Separate Peace'</ta>

<detailed_au>

<first>Walter</first>

<last>Farr</last>

<suffix>Daily Mail Special Correspondent</suffix>

```
<wd pos="547,1273,622,1298">From</wd>  
<wd pos="638,1271,784,1297">WALTER</wd>  
<wd pos="799,1272,890,1299">FARR</wd>  
<wd pos="906,1273,989,1303">(Daily</wd>  
<wd pos="1004,1273,1065,1299">Mail</wd>  
<wd pos="1080,1272,1182,1303">Special</wd>  
<wd pos="1199,1273,1412,1304">Correspondent)</wd>  
<wd pos="1446,1286,1649,1318">Washington,</wd>  
<wd pos="1668,1288,1766,1319">Friday.</wd>  
<wd pos="574,1325,802,1367">BRITAIN,</wd>  
<wd pos="831,1328,894,1362">the</wd>  
<wd pos="925,1326,1058,1362">United</wd>  
<wd pos="1087,1328,1214,1369">States,</wd>  
<wd pos="1239,1328,1366,1372">China,</wd>  
<wd pos="1394,1331,1529,1373">Russia,</wd>  
<wd pos="1553,1333,1624,1368">and</wd>
```

The screenshot displays a digital newspaper viewer interface. The main window shows a newspaper page titled "The Cabinet Newspaper." The page is overlaid with several colored rectangles: a yellow rectangle at the top, a red rectangle below it, a green rectangle below that, and a blue rectangle on the right side. The text "The Cabinet Newspaper." is visible at the top, and "per." is visible on the right side. The interface includes a top navigation bar with "Page 01" and "ERNEST". Below the navigation bar, there are fields for "Article 1 of 50", "Clip 1 of 1", "Source Page: [1]", and "Category: Frontmatter". On the right side, there is a sidebar with the following information:

- Select Issue: NCCOP0012-C00001A0000011 - Banc
- Number of pages: 8
- Issue Complete:
- Article Title: The Cabinet Newspaper
- Article Metadata:
 - Article ID: 8818-881
 - Article Type: Article
- Issue Metadata:
 - Newspaper: Cabinet Newspaper
 - Edition: Second Edition
 - Issue Number: Issue: 1
 - Vol. Num: 1
 - C_Date: November 27, 1858
 - G_Date: 1858 November 27
 - GrchDate: 18581127
 - Day: Saturday

Quality Assurance (QA) 质量保证 (QA)

The screenshot displays a digital library interface for the AIMA-ED database. The main content area shows a scanned page from the book "Literature. The Loyal Heart; or The Trappers. By GUSTAVE ALMARD. Routledge and Co., London, 1858." The page is titled "Literature" and contains a chapter titled "A GALLENT BOY". The text describes a scene in Mexico, where a crowd of people is gathered, and a young man is being pursued by a horse. The text is highlighted in yellow.

The interface includes the following elements:

- Page Info:** AIMA-ED, Report Page, Select Issue: HCCOF0012-C00001-180000011 - Desc
- Article Info:** Article: 5, Page: 30, Chap: 1, Source Page: 2, Category: Fiction, drama, Number of pages: 8, Issue Complete
- Article Title:** Literature, Sub-Title: The Loyal Heart; or the Trappers, By Gustave Almard, Routledge and Co., London, 1858
- Article Metadata:** Article ID: 00070-002, Article Type: Article
- Author:** C_Auth: Gustave Almard, Q_Auth: Gustave Almard
- Issue Metadata:**

Quality Assurance (QA) 质量保证 (QA)

The screenshot displays a library catalog interface for AIMA-ED. The main record shows the following details:

- Article: 5 of 59
- Clip: 1 of 2
- Source Page: 2
- Category: Fiction, drama
- Number of pages: 8
- Issue Complete:

The article title is "Literature. The Loyal Heart; or The Trappers. By Gustave Aimard. Routledge and Co. London, 1858." The sub-title is "The Loyal Heart; or the Trappers. By Gustave Aimard. Routledge and Co. London, 1858." The author is Gustave Aimard.

A metadata popup window is open, showing the following information:

- Article ID: 00020-002
- Article Type: Article
- Author: Gustave Aimard
- C_Auth: Gustave Aimard
- G_Auth: Gustave Aimard

The main text of the article is visible at the bottom of the screen, starting with "LITERATURE WITH his blade as all who ventured too near him, or whose unfortunate destiny presented them from getting out of his way as fast as they would have wished. "Eh! eh! Capitan!" said, as the boy brushed him in passing, a vaquero with a stupid countenance and athletic build. "Devil take the mischief, he nearly knocked me down! Eh! but," added he, after having cast a glance at the young man, "if I mistake not, that is Enfilé, the son of my neighbour! Wait a little, please!" While speaking this aside between his teeth, the vaquero unrolled the lasso which he wore fastened to his belt, and set off running in the direction of the horseman. The crowd, who understood his intention, applauded with enthusiasm. "Bravo, bravo!" cried they. "Don't miss him, Cornejo!" encouragingly shouted some vaqueros, clapping their hands. Cornejo, since we know the name of this intemperate vaquero, valued himself

Application Build (Finally)


检索举例

Home

Advanced Search

Browse by Date

Essays

Resource Links 

Daily Mail

Historical Archive 1896-2004








Churchill

Search

Search Settings

- Entire Document (every word)
- Keyword (article title and citation data)

Limit Results by Publication Date(s)

All  Date  Month  Year 
And Date  Month  Year 

For more detailed options, try [Advanced Search](#)

Try [Gale NewsVault](#) to cross-search the Daily Mail with other historical newspapers.

Why Digitize Primary Sources? 为何将原始资源数字化？

Digital Archives can bring those faraway mines right to their local libraries – the “**raw thing**” unaltered and unedited.

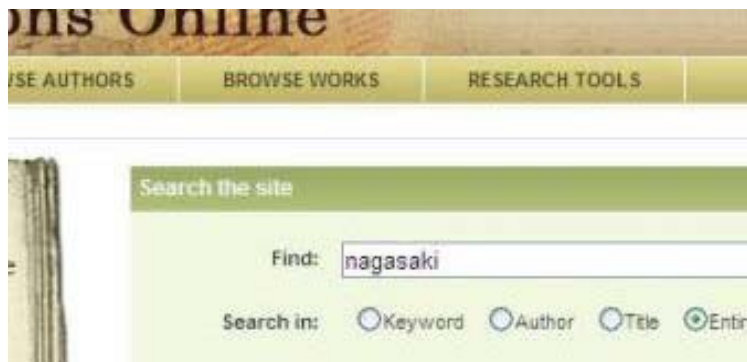
数字档案可以将远方的资源引入当地图书馆——“**原始文献**”是未经修改和编辑的。



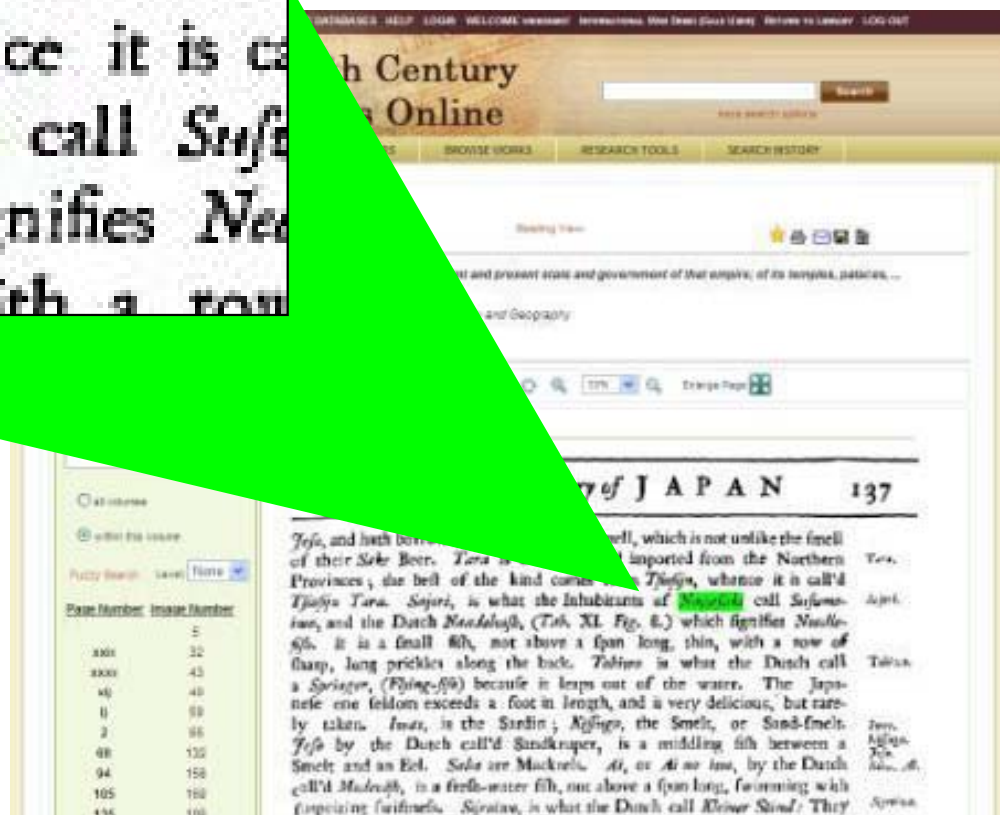
Why Digital Archives? 为何选择数字档案？

Moreover, Gale's digital archives utilizes OCR technology to allow **full-text searching** of these valuable documents.

此外，**Gale**的数字档案通过**OCR**技术使您可以对珍贵的文档进行**全文检索**。



...imported from the North
from *Tsiosya*, whence it is ca
tants of **Nagasaki** call *Syfo*
Fig. 8.) which signifies *Ned*
an long thin with a tom



Why Should Students be Exposed to Primary Sources? 为何鼓励学生使用原始资源？

History Books are Not Enough

历史类书籍是不够的

- History books are **summaries** of events neatly written by historians

历史图书是由历史学家撰写的关于事件的总结

- They are great ways of conveying knowledge in a compact form, but they allow students to be **passive** in their studying

图书是紧密地传播知识的一个很好的途径，但是他们会使学生在学习当中处于被动

- Passive studying can lead to **rote learning, uncritical swallowing** of facts, and may develop a dislike of history altogether

被动学习会导致死记硬背、不加批判地接受事实，并且可能会不喜欢学习历史

What is the difference between DA and Database? 数字档案和数据库的区别？

Database: 数据库:

- A collection of more current data/sources 更多的是当今数据/资源的合集
- Cover current events 包含当今的事件
- They get updated usually 经常更新
- Usually available for subscription only 通常是订阅模式

Digital Archive: 数字档案:

- Accumulation of historical records 历史记录的累积
- Records usually available since decades or centuries ago
记录通常为几个世纪以前或几十年前的
- They do not get updated at all time 不会随时更新
- Usually available for purchase & subscription 通常可以订阅或买断
- More 以及更多

What are the benefits from paid digital archive services? 付费数字档案资源具备哪些益处？

- **Original contents** 原始内容
- **Usually cover the complete contents** 通常收录完整的内容
- **Well indexed and organized contents** 优质的索引和有组织的内容
- **Full text searchable** 全文检索
- **Update with latest technologies and compliance with library standard**
最新的技术和符合图书馆标准
- **Full of features and functions** 丰富的功能
- **Ownership** 拥有权
- **Cross-searching** 可交叉检索
- **Quality Assurance (QA)** 质量保证
- **... many others** 以及更多

- Gold and Digital Primary Sources are “free” if you can access to them
DIRECTLY

如果您可以直接获取，金子和数字原始资源是“免费”的

- When they been crafted to become products, there are development COST
当他们被精心制作为产品，是需要开发成本的

There are many factors to determine if a product can be available in Open Web, it depends on the nature of the contents, also from legal standpoint, ownership, to costing and funding.

有很多因素可以决定一个产品可否在网络上开放获取，比如产品属性、法律角度、所有权、费用和资金

Thank you

谢谢

欢迎您参与上机培训：

5月15日 上午 09:50-10:30

电子信息与电气工程学院 4号楼

第二培训教室（208房间）