

ScienceDirect
Backfiles and Archiving
Sao Paulo
September 19, 2007

Presented by:

Lindi Belfield – Senior Product Manager



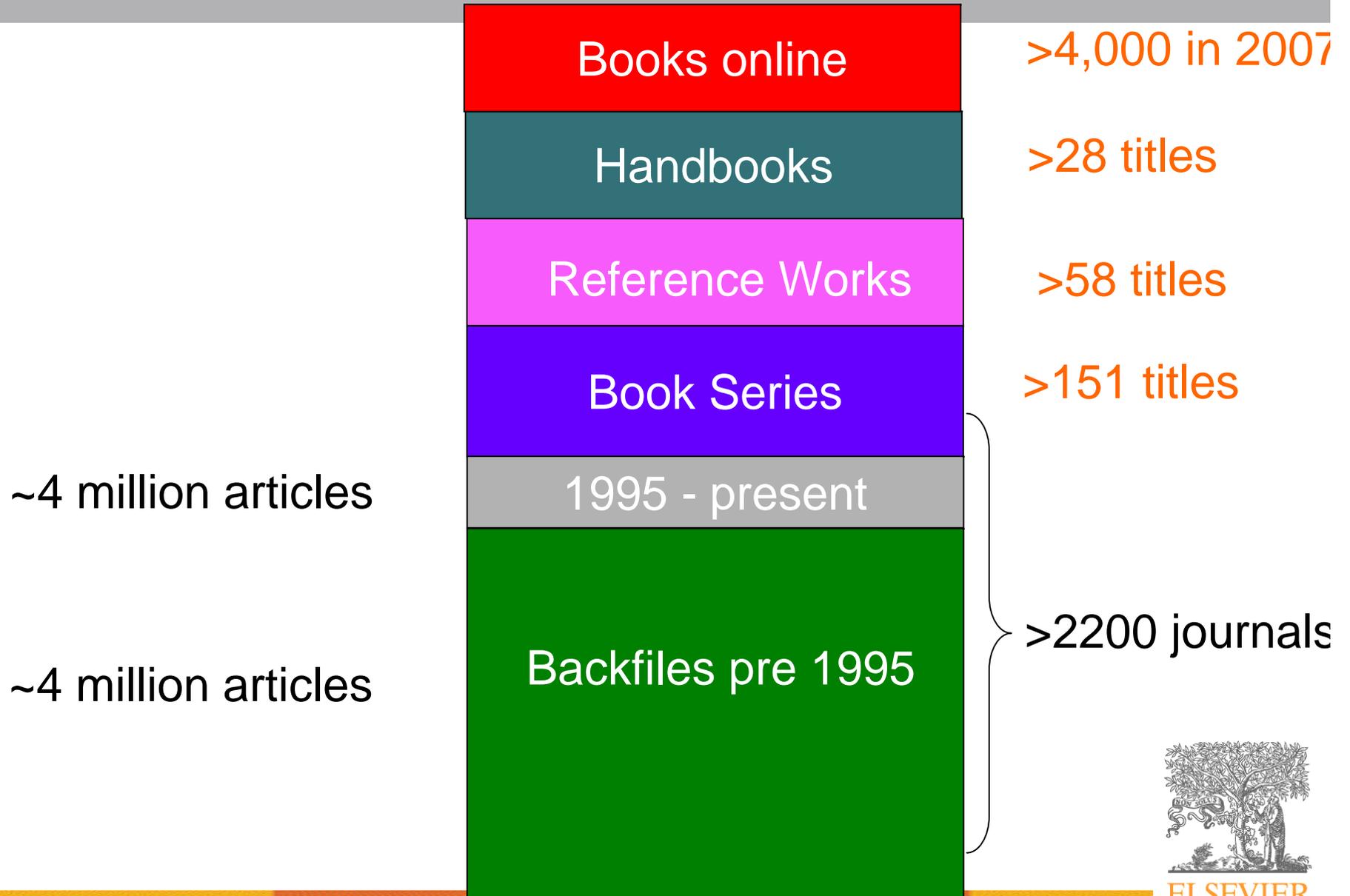
Agenda

- ScienceDirect content
- Sourcing and digitization
- Archiving and Preservation Policy
- “The Engine Room”



ELSEVIER

Growing ScienceDirect Content





Back to the
future with.....

Back Files!!!!

Back Files Project on ScienceDirect

- Elsevier's initiative to digitise all Elsevier owned journals from 1994 back to volume 1 number 1
 - Started in 2000 and will be completed 2006
 - 1800 titles + discontinued, changed and split titles
 - Compiled and sold as complete subject packages



Four Sea Containers and Two Air Containers



Millions of pages
\$40 million for scanning
4 million articles < 1995

A New Factory – Just for Backfile Production

2000 people scanning and processing the articles



ELSEVIER

Rows of Storage



Chemistry

Earth Sciences

Pharmacology

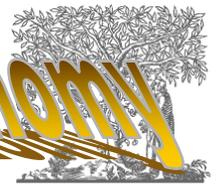
So far..... 29 backfile packages
launched. A major milestone
is the entire file for

The Lancet

Economics

Biochemistry

Physics and Astronomy

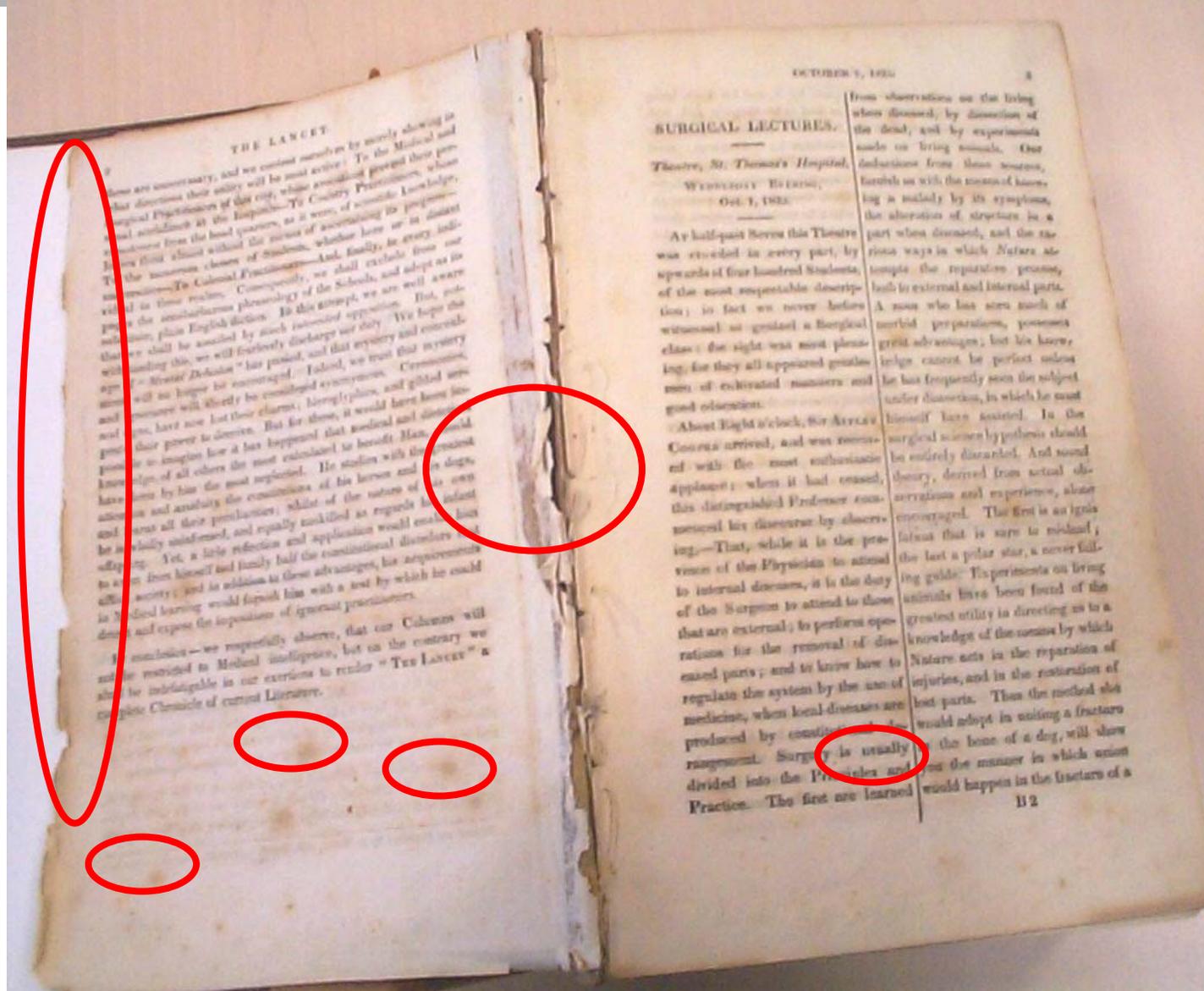


ELSEVIER

Material Sciences

Social Sciences

Lancet – Volume 1, Number 1, 1823



Special techniques to improve old page images

OCTOBER 5, 1823.

3

SURGICAL LECTURES.

Theatre, St. Thomas's Hospital,
WEDNESDAY EVENING,
Oct. 1, 1823.

At half-past Seven this Theatre was crowded in every part, by upwards of four hundred Students, of the most respectable description; in fact we never before witnessed so genteel a Surgical class: the sight was most pleasing, for they all appeared gentlemen of cultivated manners and good education.

About Eight o'clock, Sir ASTLEY COOPER arrived, and was received with the most enthusiastic applause; when it had ceased, this distinguished Professor commenced his discourse by observing,—That, while it is the province of the Physician to attend to internal diseases, it is the duty of the Surgeon to attend to those that are external; to perform operations for the removal of diseased parts; and to know how to regulate the system by the use of medicine, when local diseases are produced by constitutional derangement. Surgery is usually divided into the Principles and Practice. The first are learned

from observations on the living when diseased, by dissection of the dead, and by experiments made on living animals. Our deductions from these sources, furnish us with the means of knowing a malady by its symptoms, the alteration of structure in a part when diseased, and the various ways in which Nature attempts the reparative process, both to external and internal parts. A man who has seen much of morbid preparations, possesses great advantages; but his knowledge cannot be perfect unless he has frequently seen the subject under dissection, in which he must himself have assisted. In the surgical science hypothesis should be entirely discarded. And sound theory, derived from actual observations and experience, alone encouraged. The first is an ignis fatuus that is sure to mislead; the last a polar star, a never failing guide. Experiments on living animals have been found of the greatest utility in directing us to a knowledge of the means by which Nature acts in the reparation of injuries, and in the restoration of lost parts. Thus the method she would adopt in uniting a fracture in the bone of a dog, will show you the manner in which union would happen in the fracture of a

B 2



ELSEVIER

Key Benefits

- Desktop access to a large pool of research articles and all the Elsevier journals' abstracts – 24x7x365
- Access to historical data dating back to e.g., The Lancet-1823
- One time purchase – no recurring fees
- Lighten physical archive by tons of paper to go to - only (kilometres of shelf space !)
- Fully searchable and CrossRef links
- Free access to > 7 million abstracts
- The answer, or contingency plan in case disaster hits.....



ELSEVIER

National Bosnian Library: the rubble is books & journals

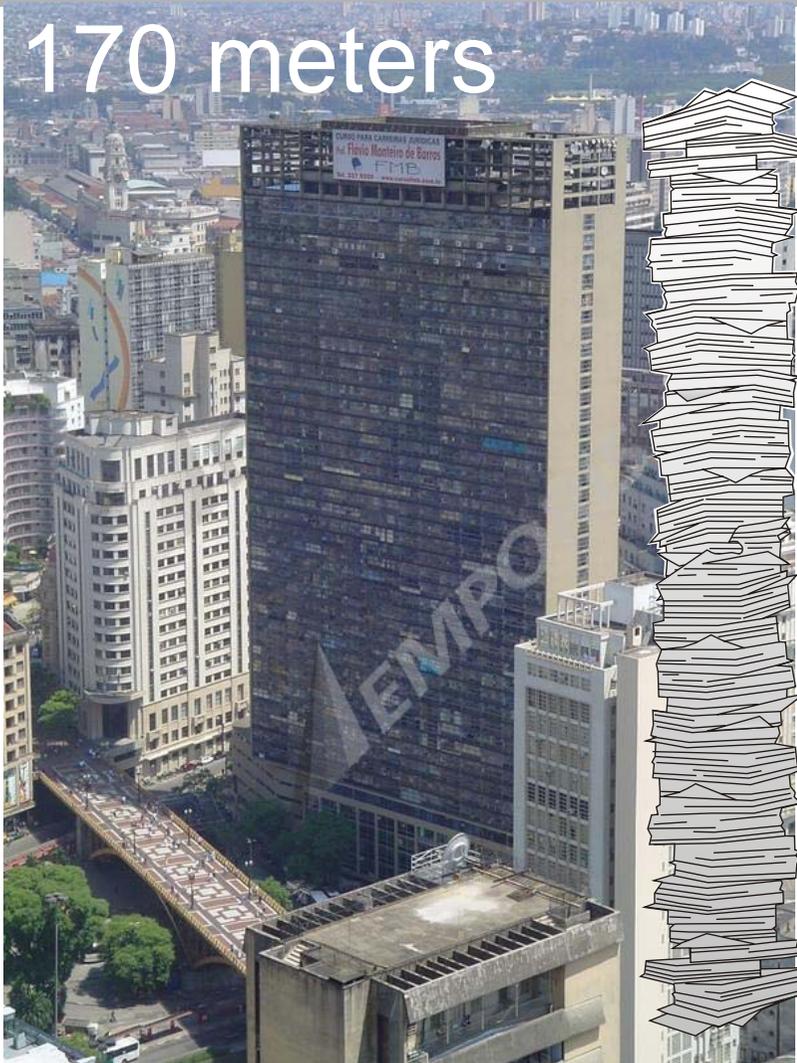


Kobe National Library - earthquake



What does 8 million articles look like?

170 meters



8 million articles

=

70 million pages

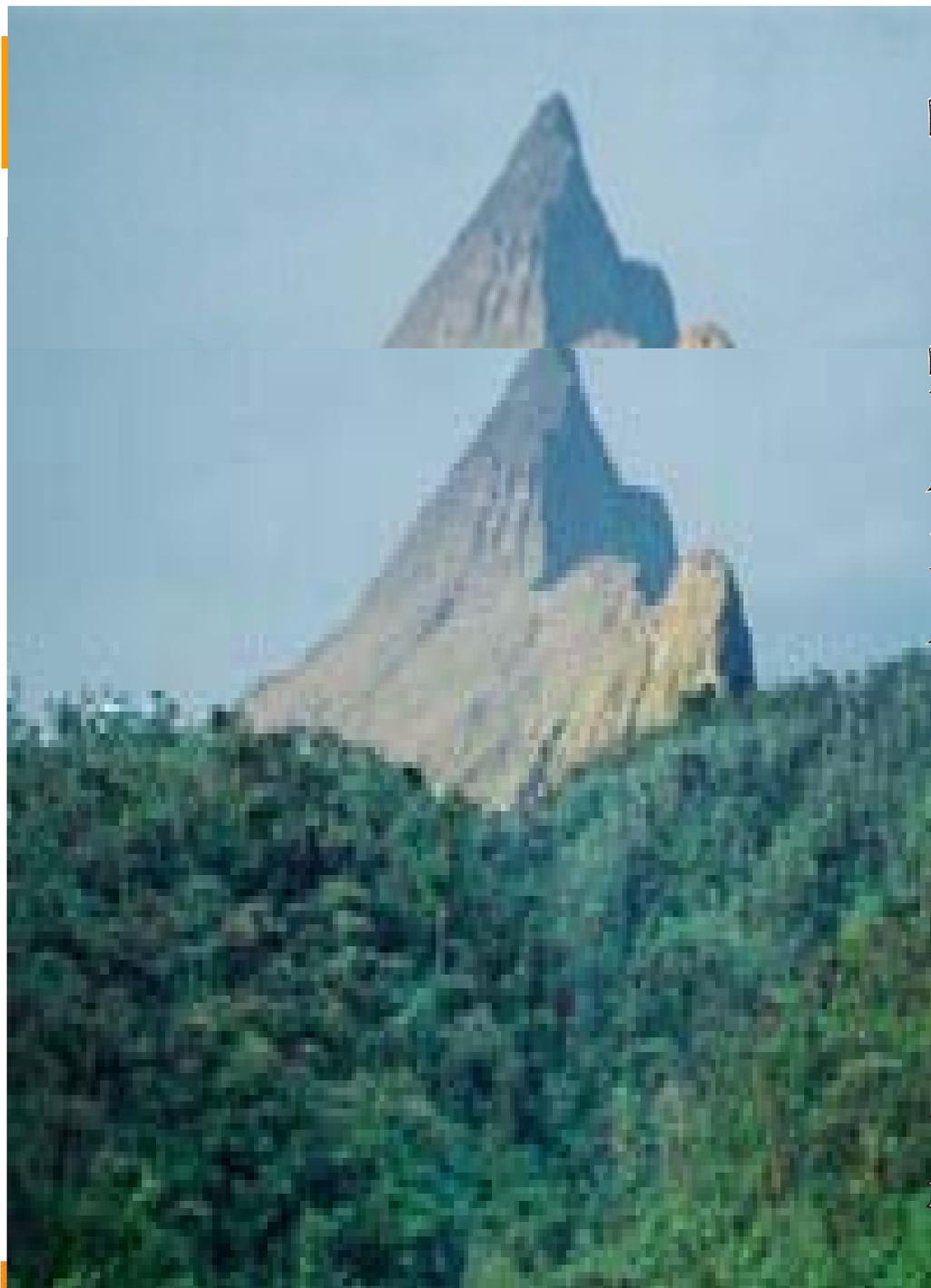
=

4 kilometers

= 24 Mirante do Vale buildings



ELSEVIER



8 million articles
=
70 million pages
=
**35 million
doublesided
sheets**
=
~4 kilometers



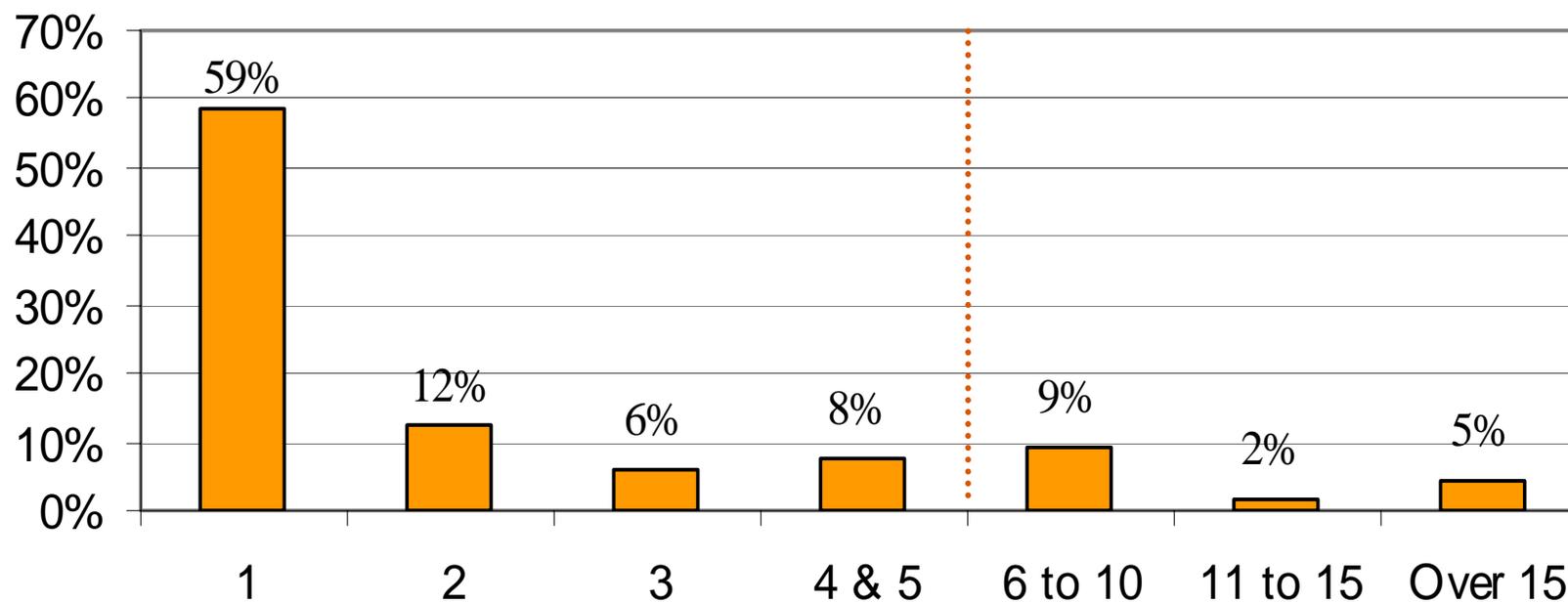
ELSEVIER

BackFiles: Value Proposition



Average usage on historical information

Percentage of researchers' readings by age of scientific article (years since publication)



16% of the articles used is older than 5 years

Source: King et al. 1998.



ELSEVIER

Importance of older research articles

Average importance rating	younger than 1 year	older than 1 year
Research purpose	4.9	5.3
Teaching purpose	4.8	5.0
Average time spent reading	43 min	61 min

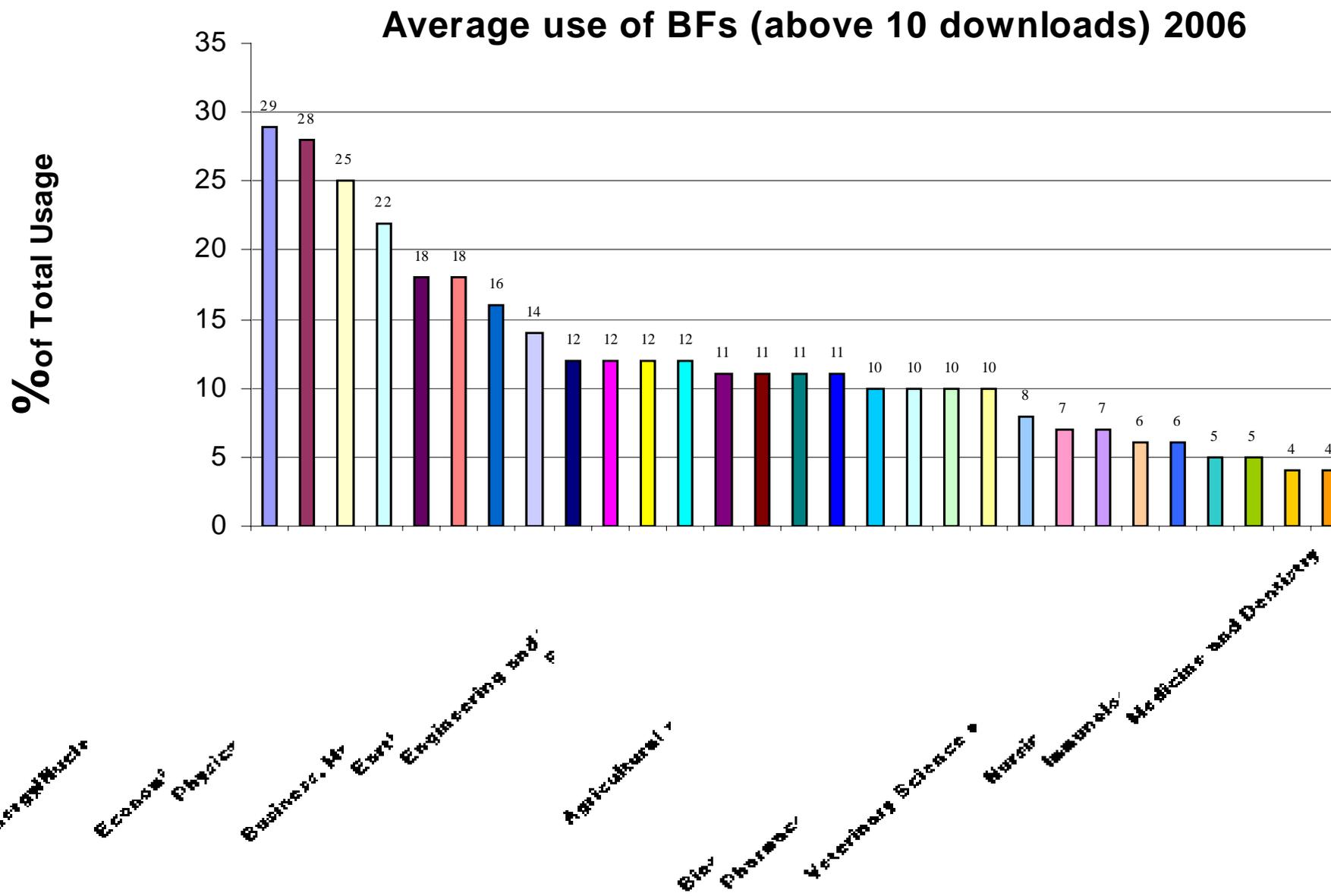
(1= weak, 7 = strong)

Conclusion is that older articles are often read for research purposes; newer articles may be read for casual interest.

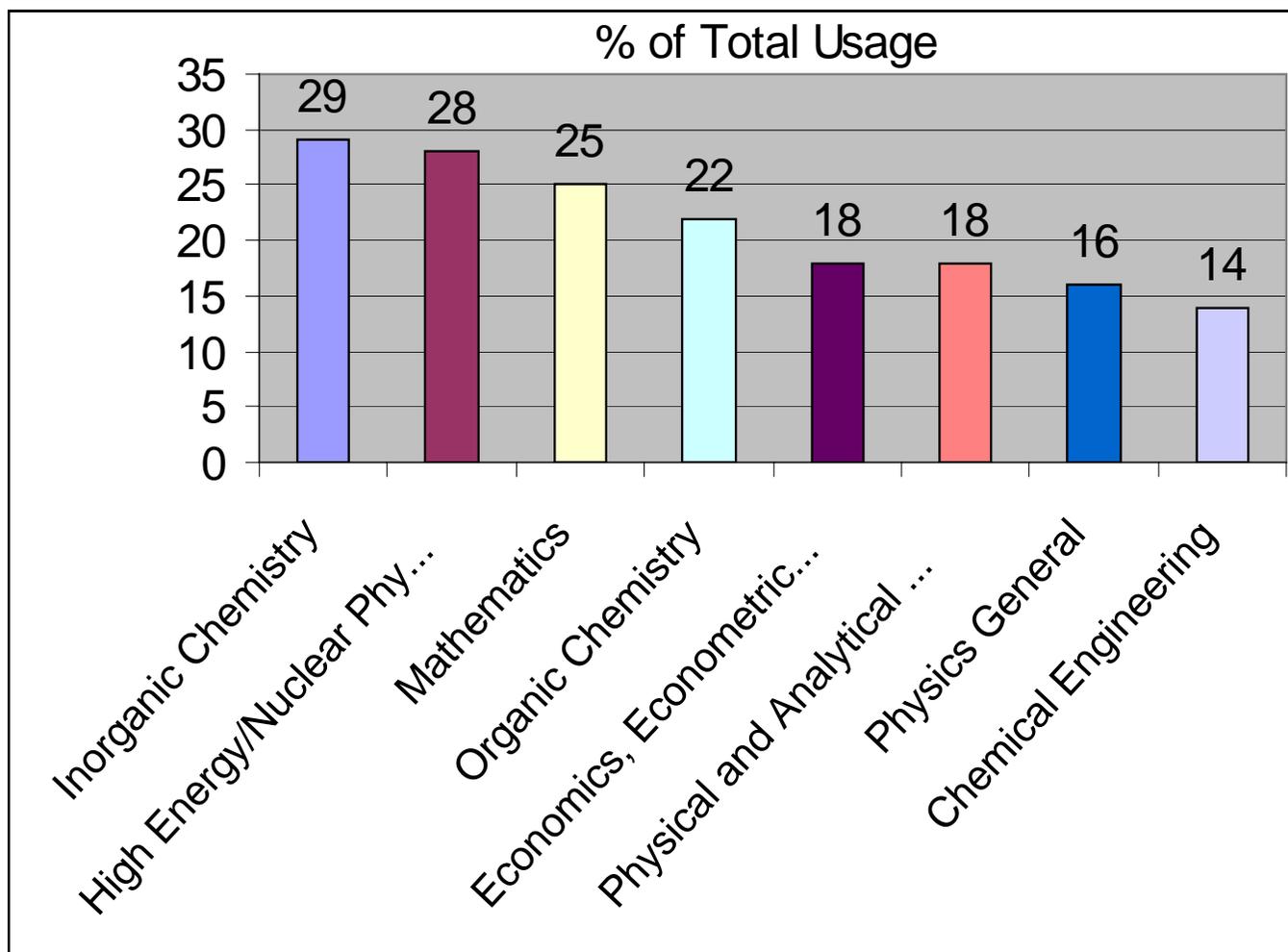
Source: King et al. 1998.



Average Usage of Backfiles on ScienceDirect 2006 = 12%

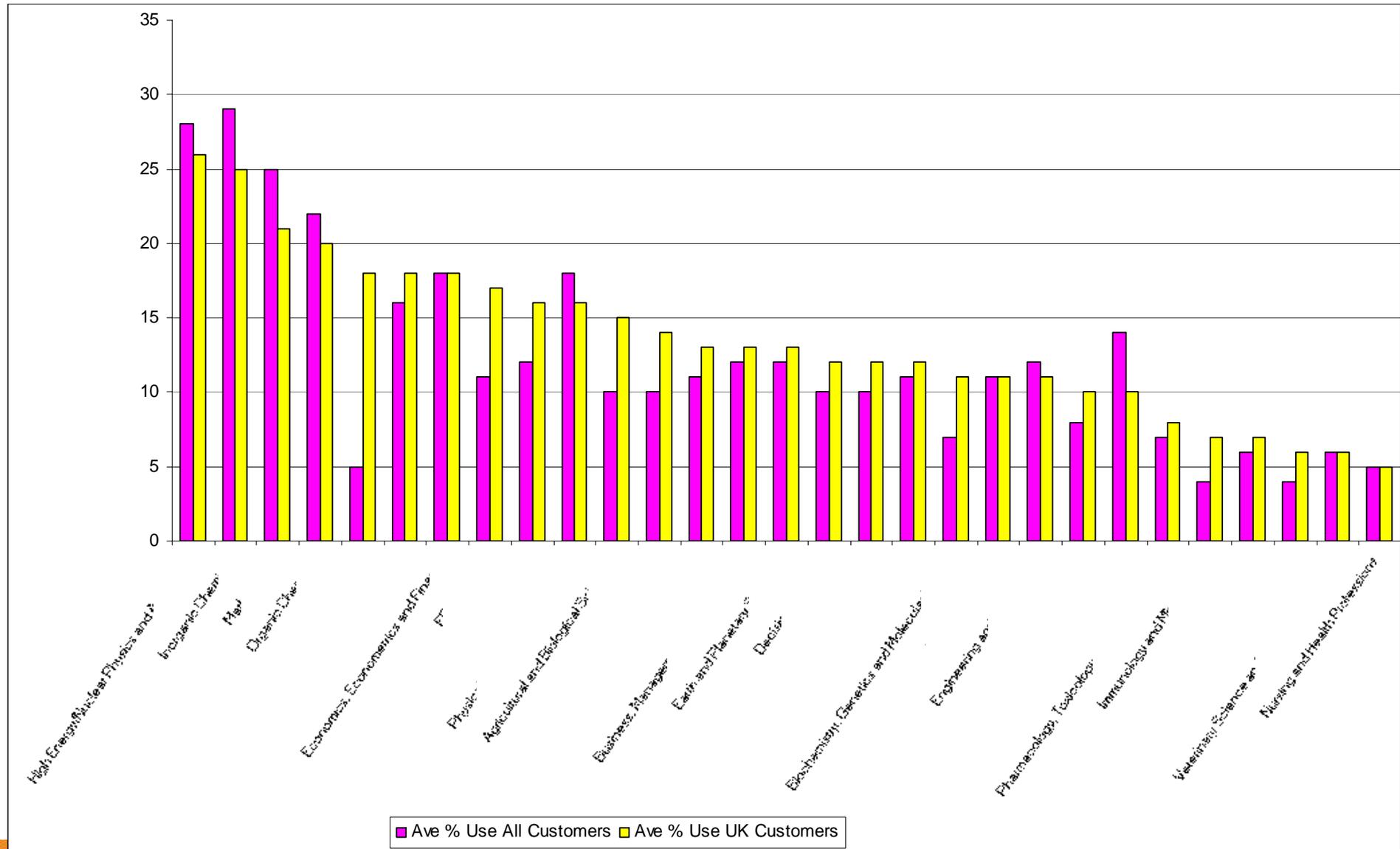


Subjects with more than average 12% use of pre-1995 backfiles



Average use of pre -1995 content is ~12% of all backfiles

Average use in the UK versus Global 2006



Overall conclusion

- Content more than 10 years old is extremely useful and valuable to readers and should continue to be made available from the desk top in the electronic world.

Source: King et al. 1998.



Archiving and Preservation



Importance of Archiving to Elsevier

- STM Publishers RESPONSIBILITY to Science - **“maintain the minutes of Science”**.
- Authors expect their research to be always AVAILABLE.
- An important OBLIGATION that has become critical to libraries - to preserve a PERMANENT record of scholarship.

History: The beginning of Digital Archiving at Elsevier

- **Since 1999, all ScienceDirect online licenses have contained an annex specifying that:**
 - 1) we will maintain a permanent digital archive of journals we own & offer on SD
 - 2) we will migrate the archive as technology used for storage or access changes
 - 3) we will transfer the archive to an independent, librarian-approved depository if we cannot maintain it.



Current Status of Digital Archiving at Elsevier

Four levels; 2 groups

a) Elsevier/publisher maintained:

- In-house systems and backups
- ScienceDirect Platform and backup at LexisNexis

b) Library or Consortia maintained:

- Local ScienceDirect On Site Loaders (ex: OhioLink, Univ. Toronto) but no formal agreement
- Formal Digital Archives



Level 1--Inhouse Digital Repository

- **Internal Repository -- Electronic Warehouse**
 - Final home for all processed journal articles, books, and backfiles
 - Located in Oxford at Elsevier's Data Center
 - Database for creating files for transmission to ScienceDirect as well as Local sites (ScienceDirect On Site)
 - Database backup procedures maintained.



Level 2: ScienceDirect Platform

- Located at LexisNexis in Dayton, Ohio
 - Building is bomb-proof and restricted access
- Copy maintained on the LN platform
- Working version of ScienceDirect
- Backup maintained geographically apart from Dayton, Ohio

Level 3: Full Set Libraries/Consortia

- De facto library archives
- ScienceDirect OnSite customers and library consortia worldwide, including about a dozen who receive one copy of every journal (ex: OhioLink, Toronto)
- No formal archival commitment, but these libraries have all the journals
- Discussions with National Archives -- special agreements with respect to SDOS licenses for those countries in which there is a desire to hold an archive “on our soil”
 - Ex: Australia, Korea



ELSEVIER

Level 4—Official Archives

- Official archives -- libraries with whom we establish a contractual relationship to preserve the digital journals in perpetuity and make them available to anyone internationally at such time as we (or a successor) no longer can do so
 - The Royal Library of The Netherlands
 - Portico (US)
 - Working closely with CLOCKSS



Official Archives: KB and Portico

- On August 20, 2002 we announced the first official ES archive: the Koninklijke Bibliotheek, which is the National Library of the Netherlands and is known as the KB
- The KB has been an outstanding leader in digital preservation activities
- They have received all files, including backfiles (v.1,no. 1) as available

Portico is our second official archive

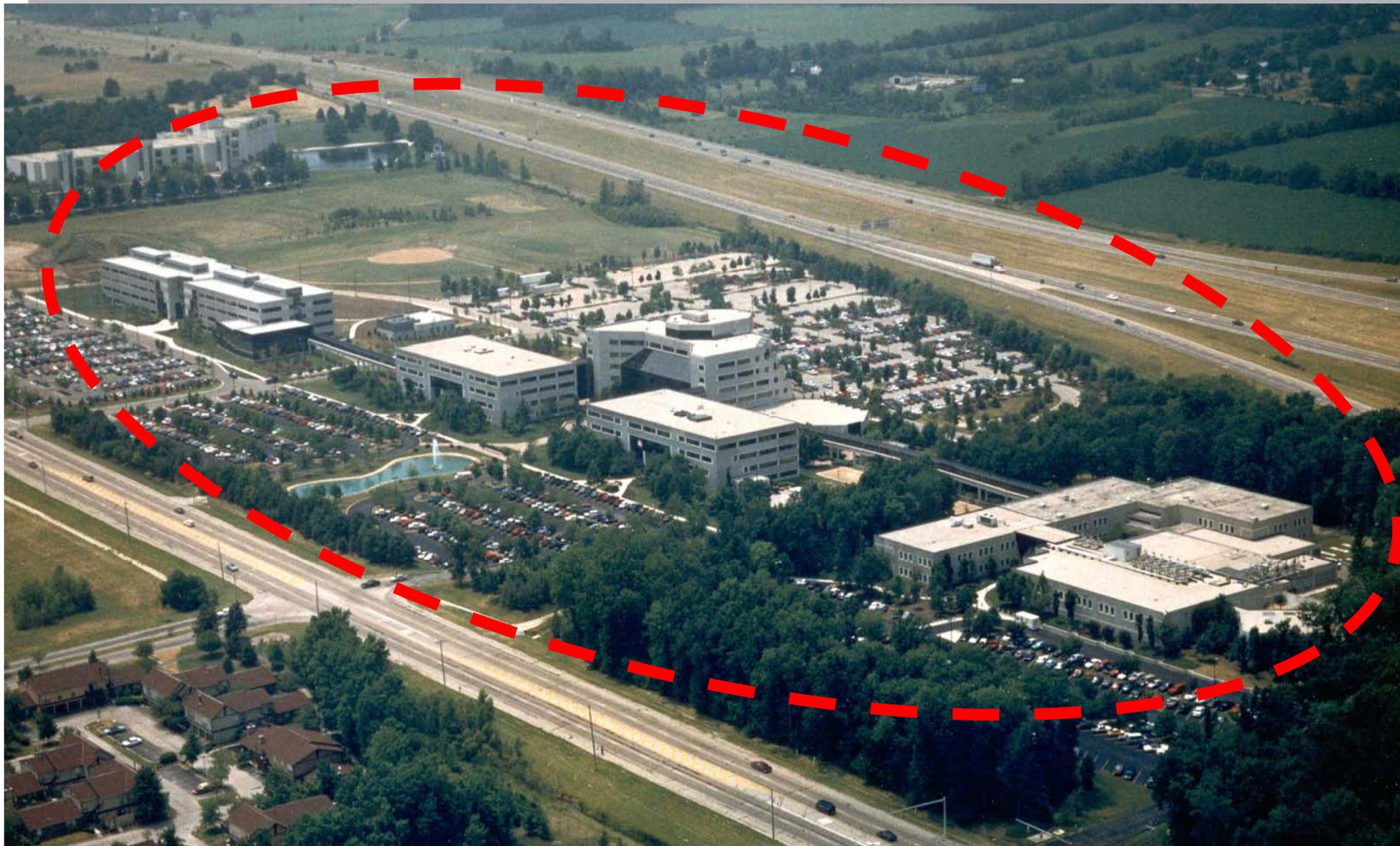
- The Portico Model also meets the requirements:
 - advanced technology;
 - experience with digitization;
 - sustainable business plan;
 - experience in the area.
 - ability to provide content online to subscribing libraries

Many people publish on the web and say it's
very cheap and easy.....

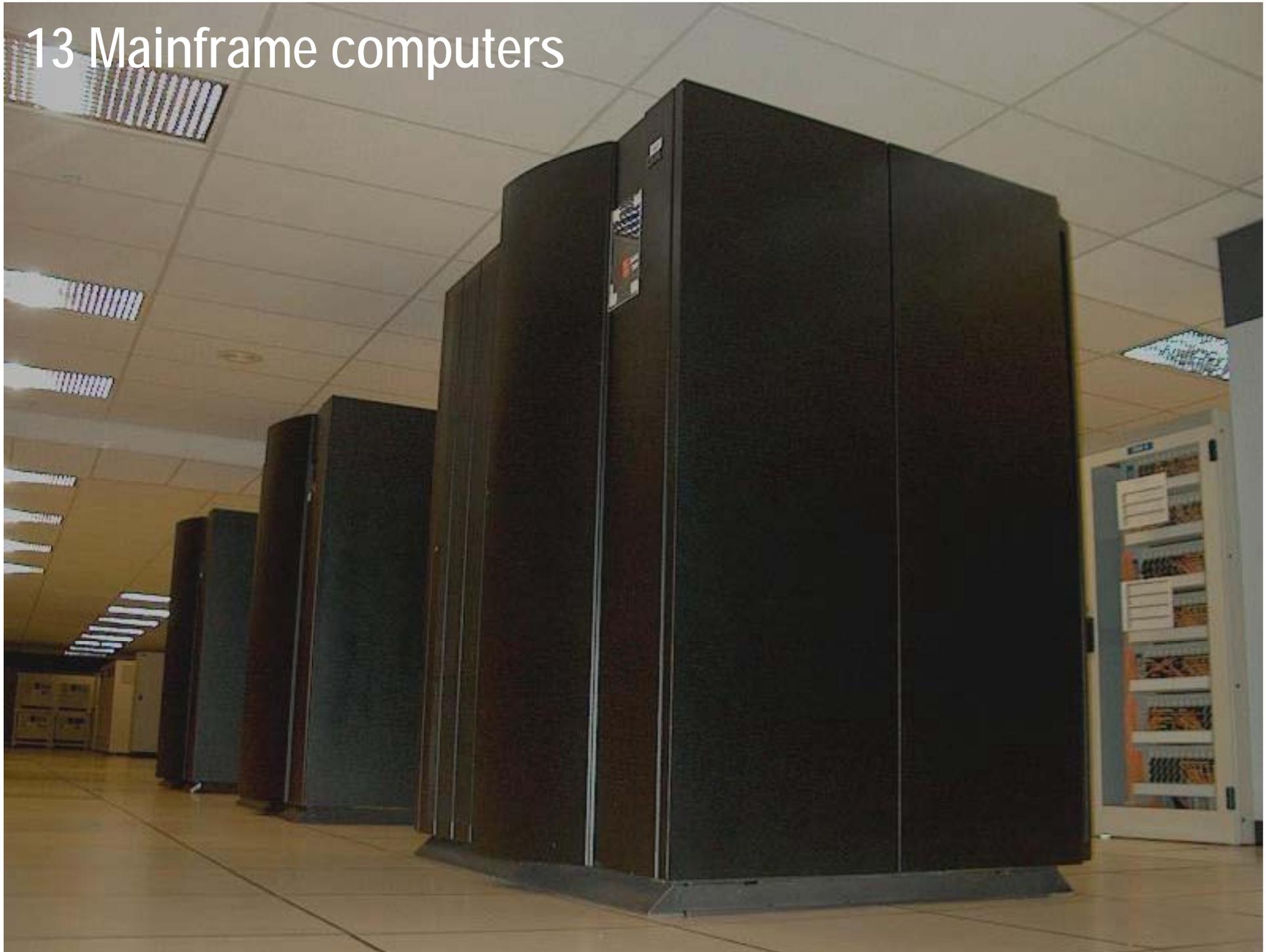
So, was ScienceDirect built easily and
cheaply in a garage perhaps.....?



Dayton Campus Ohio USA – Major Operations Site & Development



13 Mainframe computers



Rows of Servers



Storage in Petabytes!

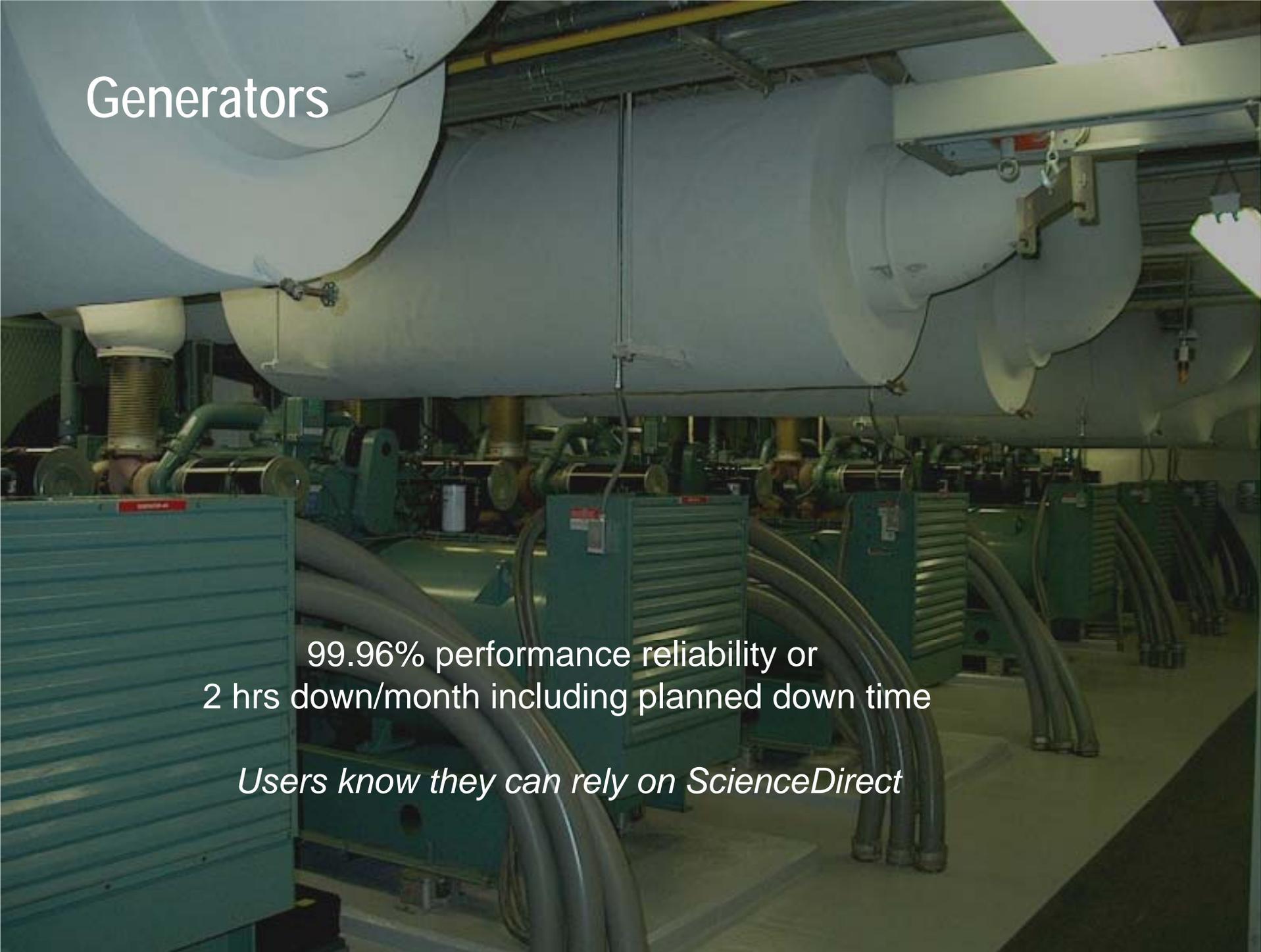


Chilled Water Plant

Fills a swimming pool every 8 minutes



Generators

A photograph of a large industrial generator room. The room is filled with rows of green generators. Large white overhead structures, possibly part of the cooling or ventilation system, are suspended from the ceiling. The lighting is somewhat dim, and the overall atmosphere is industrial and functional.

99.96% performance reliability or
2 hrs down/month including planned down time

Users know they can rely on ScienceDirect



Wet cell batteries – just in case.....

Thank You

Any Questions?

