

Revolutions in publishing

Ian Mulvany - Head of Technology - eLife

@IanMulvany

Caveats

Before going into these topics I want to set out my stall and be very clear about my biases:

Caveats

Before going into these topics I want to set out my stall and be very clear about my biases:

advocate for open access

Caveats

Before going into these topics I want to set out my stall and be very clear about my biases:

advocate for open access

no formal background in XML

Caveats

Before going into these topics I want to set out my stall and be very clear about my biases:

advocate for open access

no formal background in XML

a fan of the open web

Caveats

Before going into these topics I want to set out my stall and be very clear about my biases:

advocate for open access

no formal background in XML

a fan of the open web

some experience of building publishing workflows from scratch

Caveats

Before going into these topics I want to set out my stall and be very clear about my biases:

advocate for open access

no formal background in XML

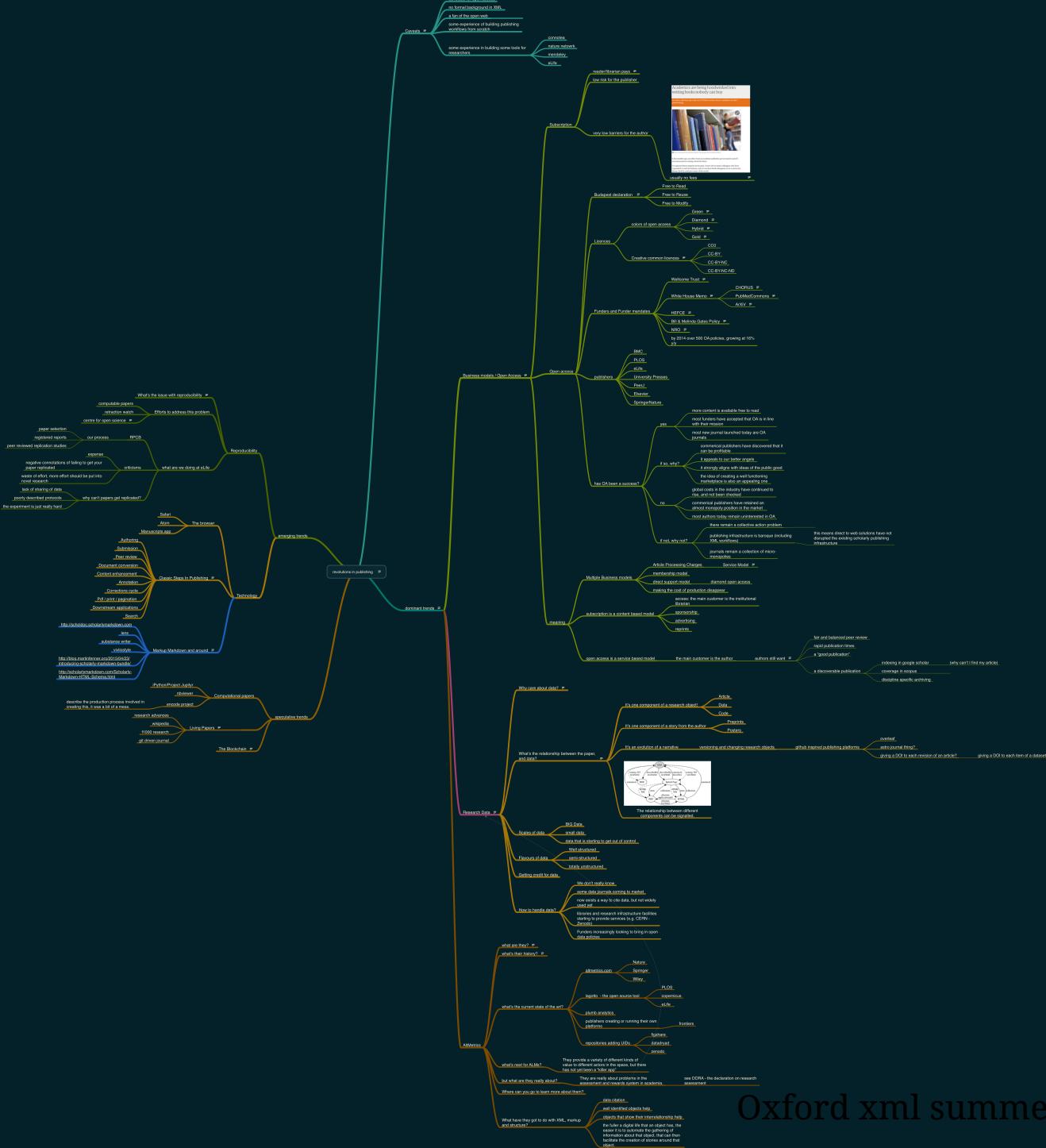
a fan of the open web

some experience of building publishing workflows from scratch

some experience in building some tools for researchers

- connotea
- nature network
- mendeley
- eLife

introduction



Oxford xml summer school 2015

- Open Access
 - Research Data
 - Object level/Alt Metrics
-
- Reproducibility
 - Webkit and the browser
 - HTML5/Markdown/JSON
-
- Computational papers
 - Living papers
 - The blockchain

dominant trends

Open Access

Research Data

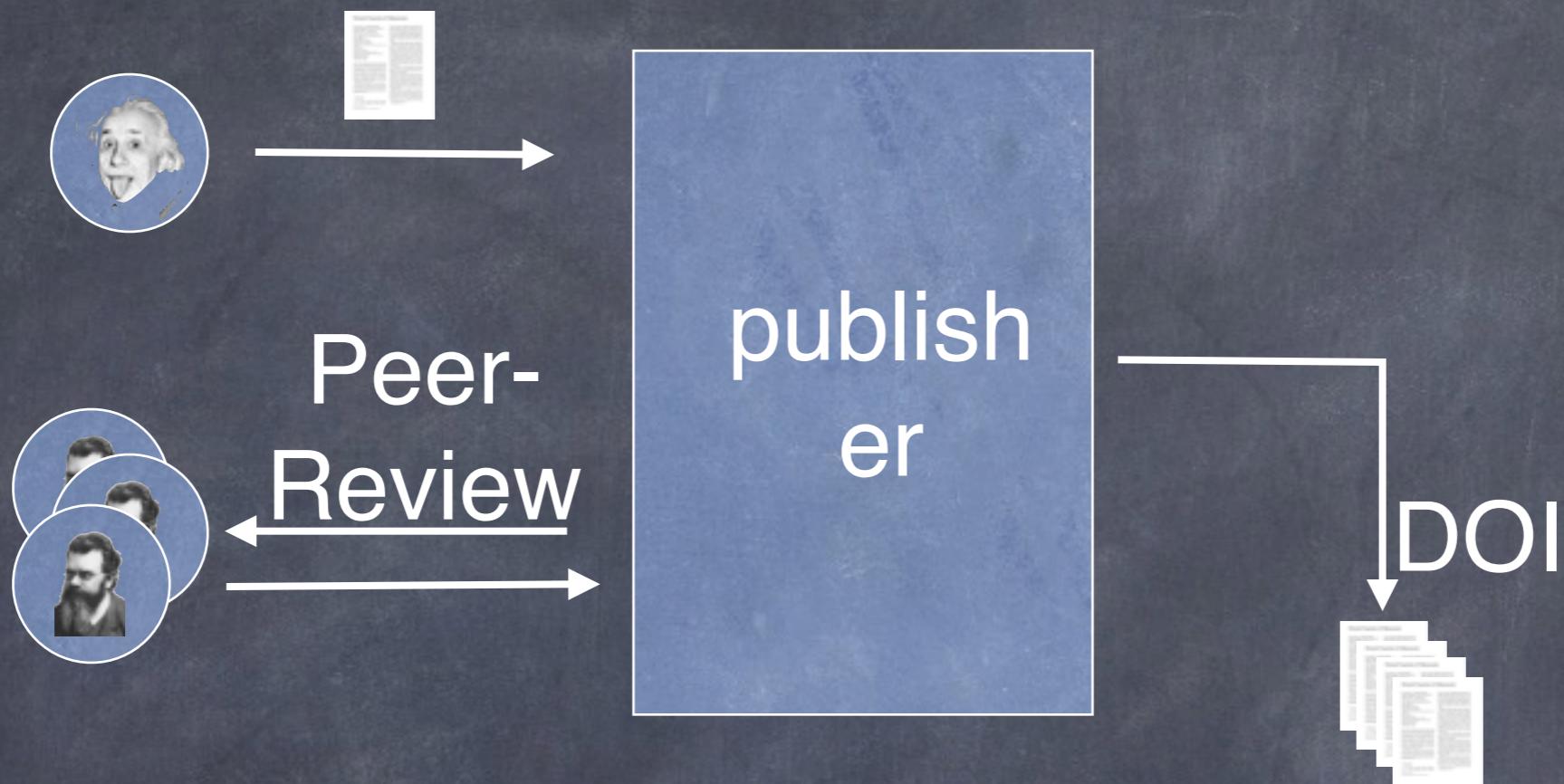
AltMetrics

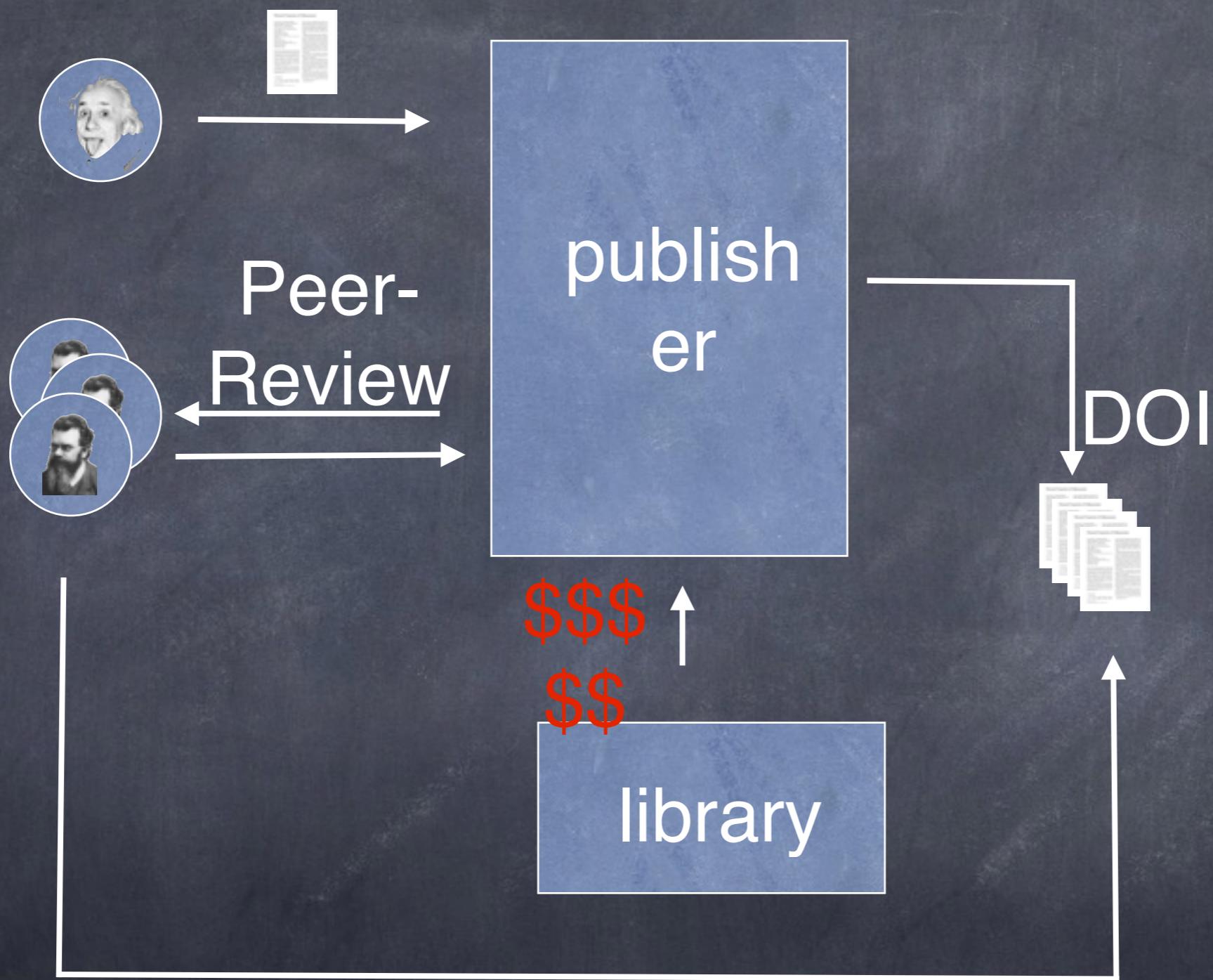
Oxford xml summer school 2015

Open Access

open access

Oxford xml summer school 2015



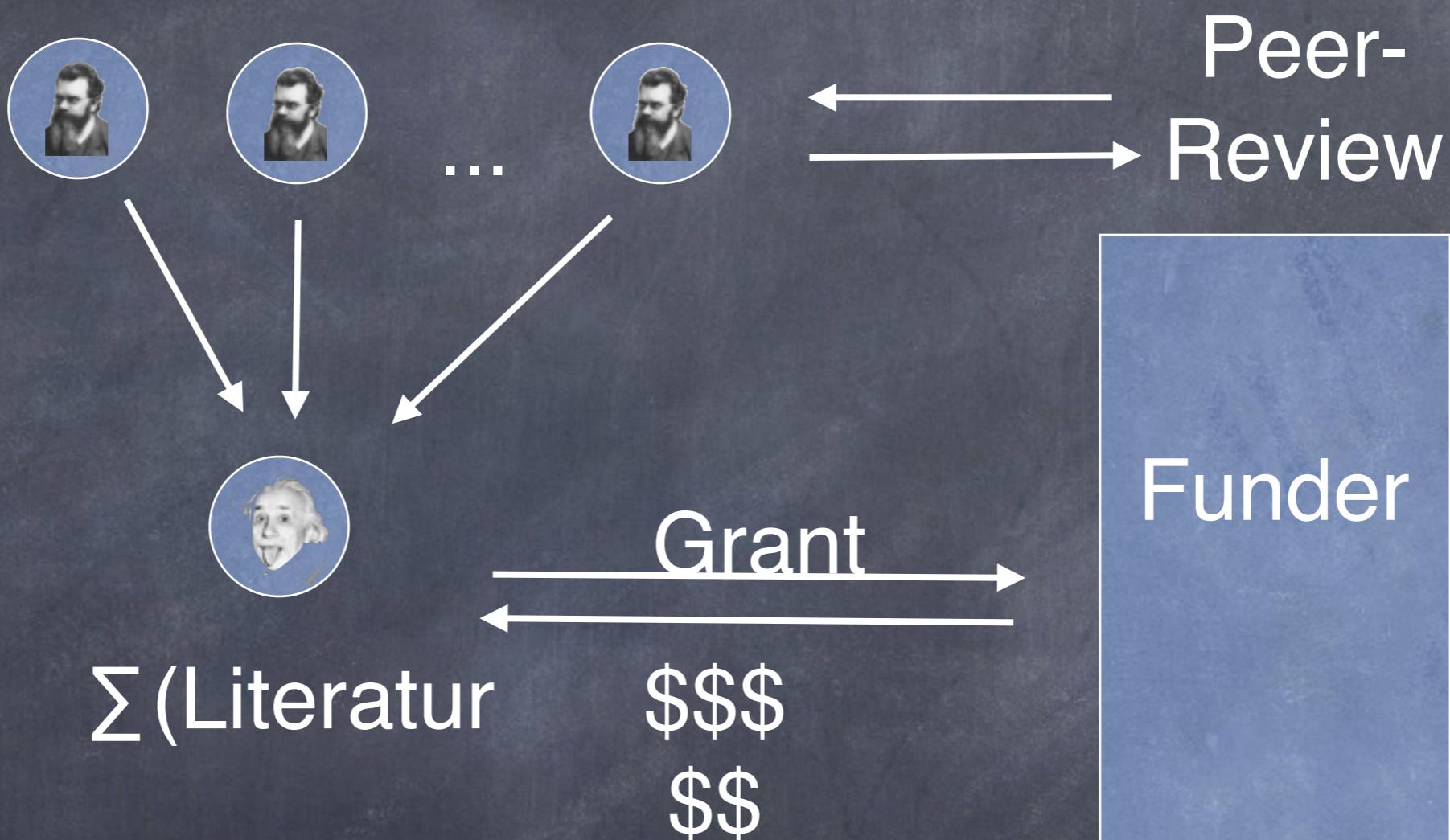


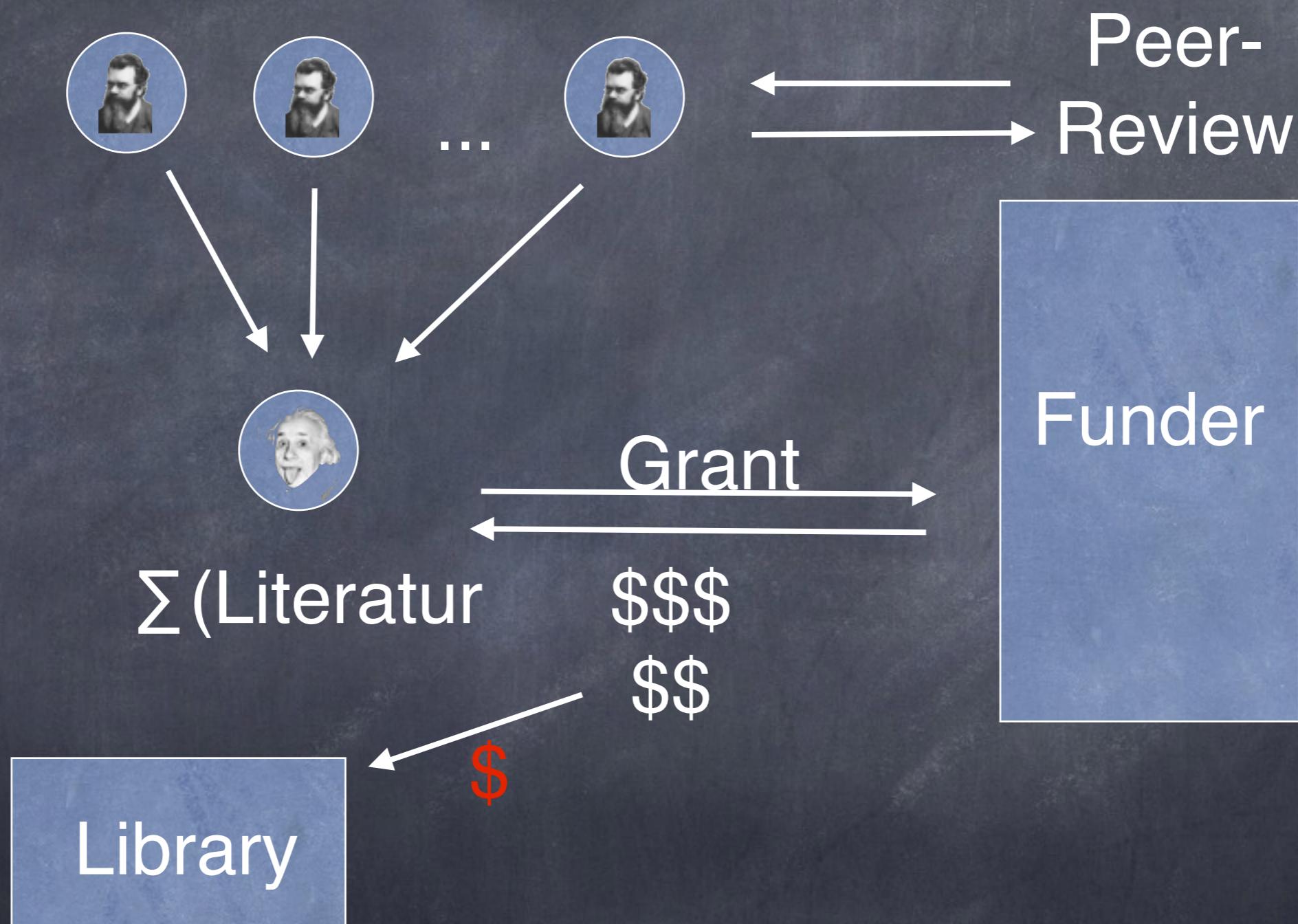
\$\$\$\$\$

Grew 3% im 2012

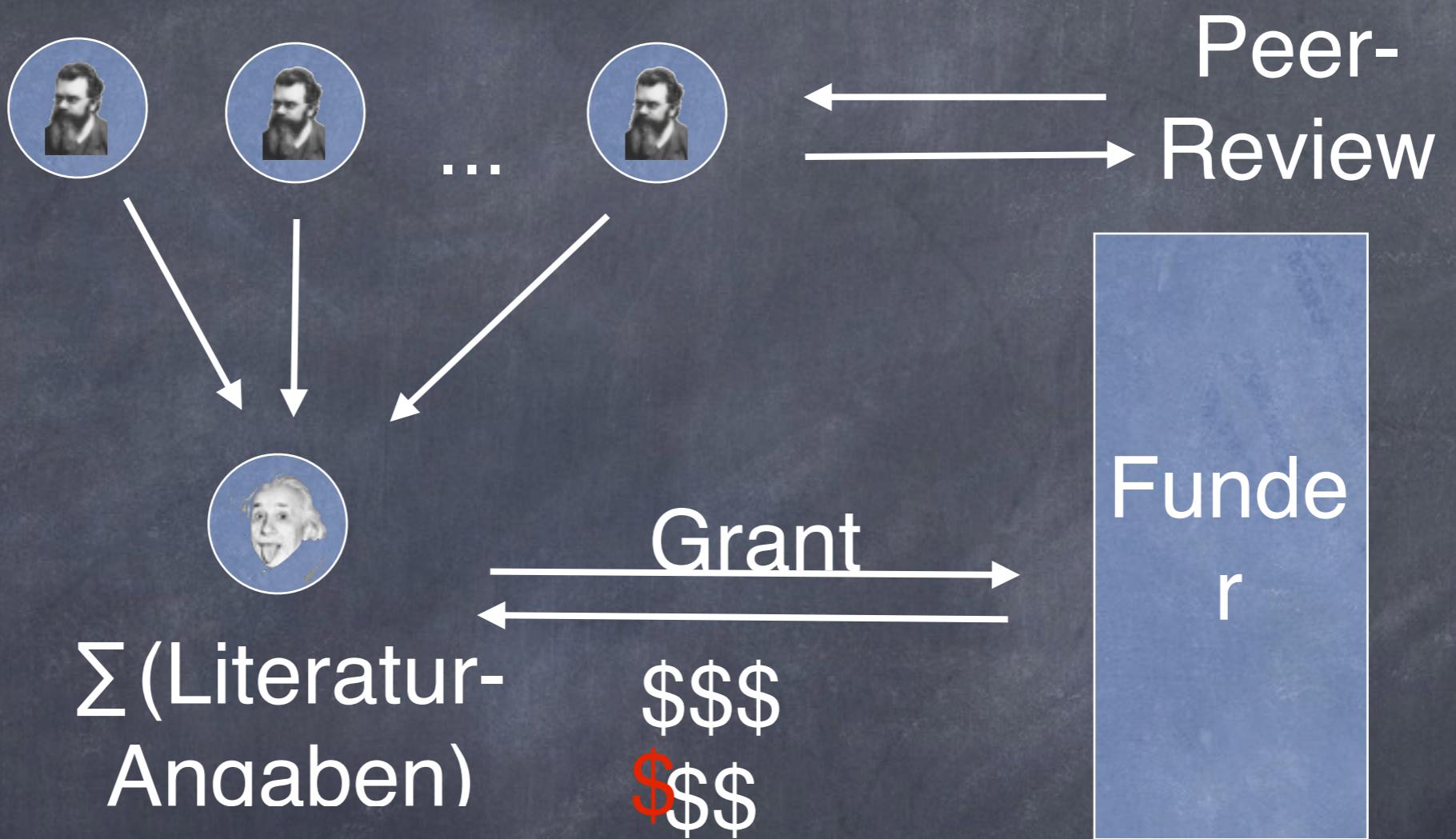
2010 \$8.9B World Net Value

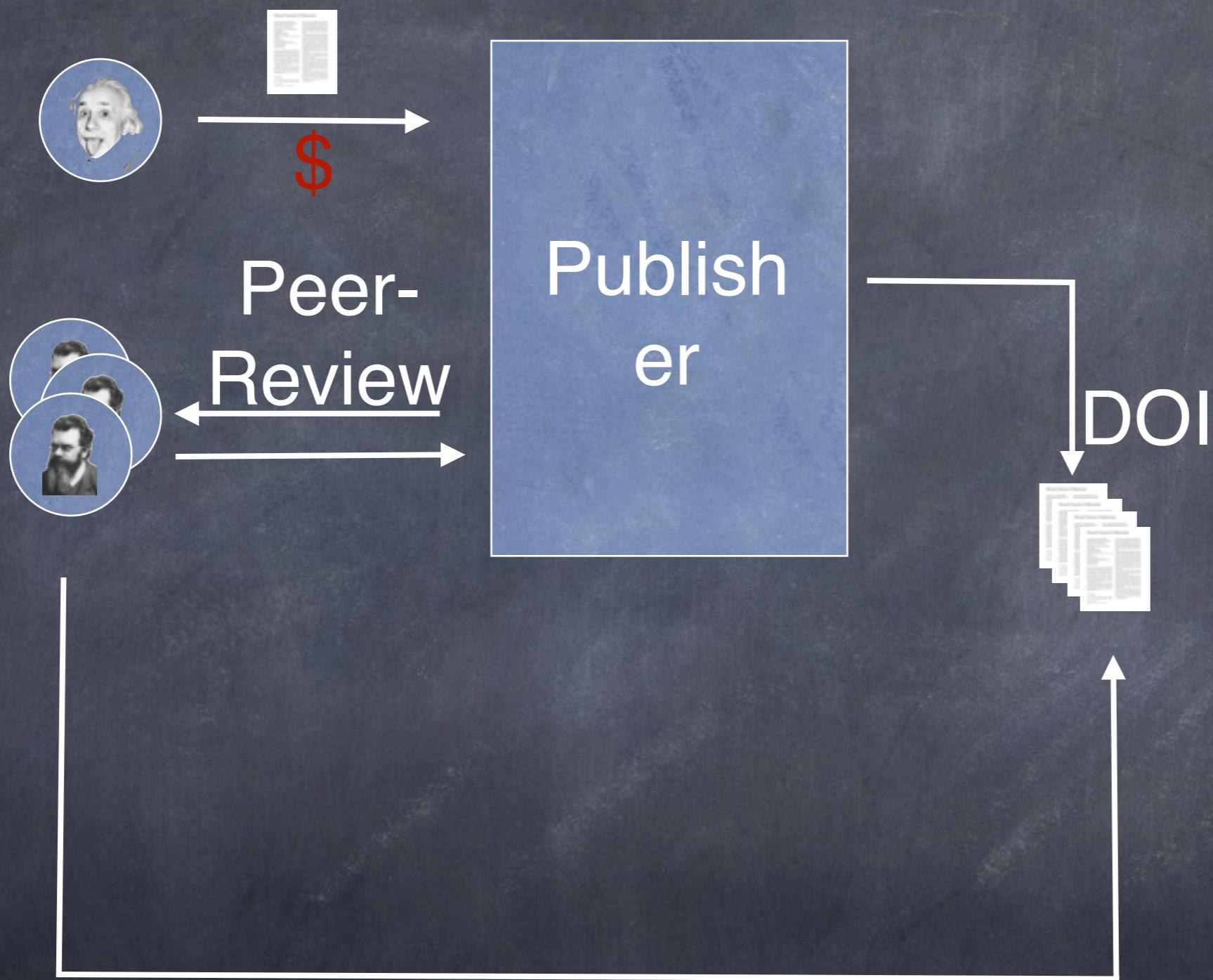
Elsevier, 2010 \$1B Profit on \$3.1B
turnover



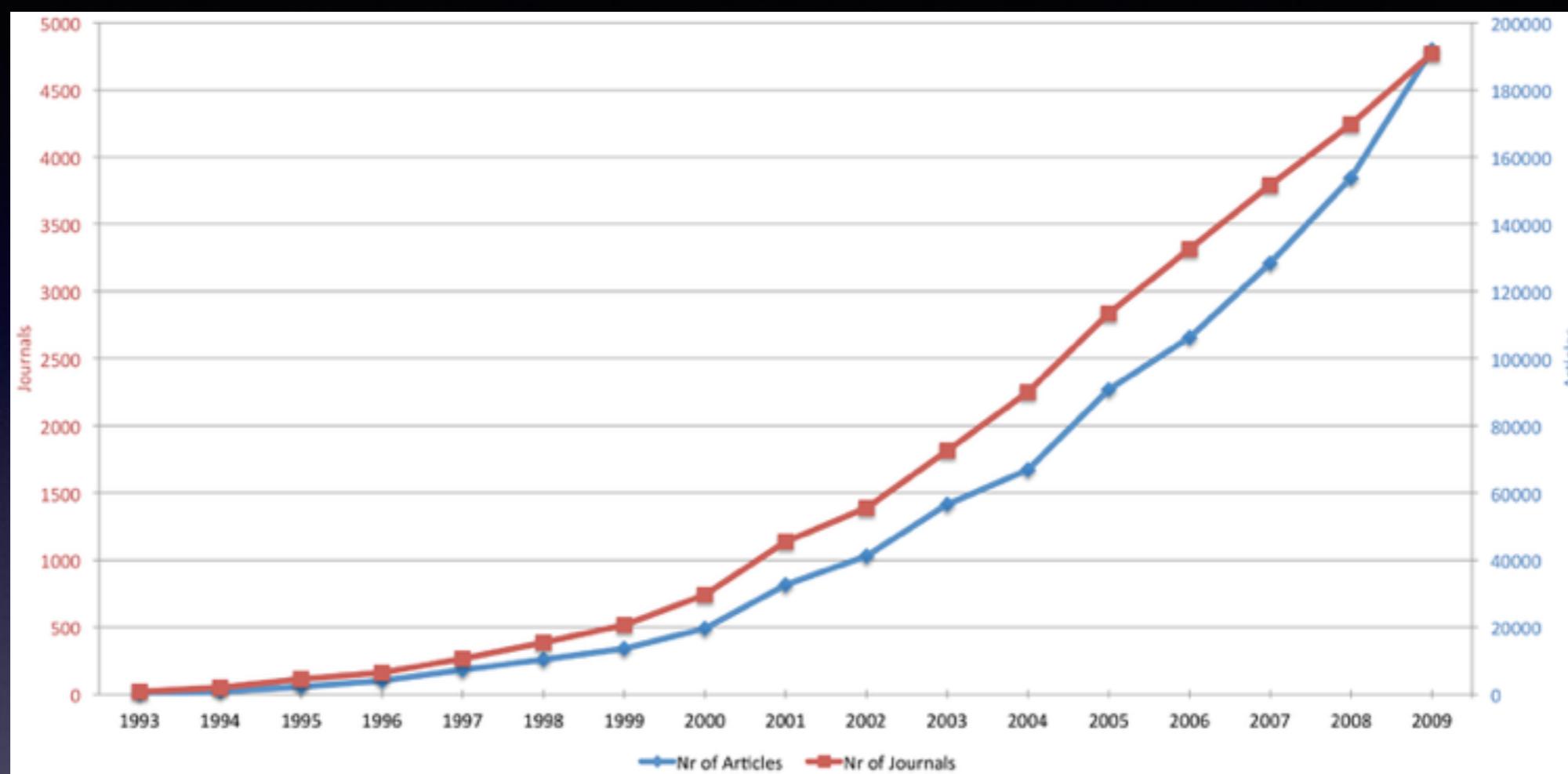


The Open Access Model





Lesen, Literaturangabe?



Subscription Benefits

- reader/librarian pays Can be argued that this fairly places the burden of the costs of production on the target market, however IMO the bundling that happens in big deals makes this argument less watertight.
- low risk for the publisher
- very low barriers for the author
- usually no fees

Open Access History

- Budapest declaration
 - February 2012
 - Meeting funded by the Soros foundation
 - 13 original signatories
- definition of open access
 - Free to Read
 - Free to Reuse
 - Free to Modify

At that time no specific license was mentioned.

(George Soros was a student of Karl Popper, and later a merchant banker)

colors of open access

- Green authors can deposit their content into a repository, some restrictions may apply around format and rights for re-use.
 - simplest to implement
 - good support for institutional repositories
 - has the backing of CAS
 - does not provide long term financial sustainability
 - under legal threat from Elsevier
- Gold
 - An article processing fee is paid for the article to be published.
 - Scales with submissions
 - Proven to work
 - Can lead to charges of predatory publishing
 - this argument is baseless
 - Applied without waivers and it becomes a barrier
 - The natural price is not yet known

colors of open access

- Hybrid
 - A journal supports a mixture of subscription based articles and gold open access
 - can transition an existing journal gradually
 - has tended to be a bit of a clusterfuck
 - leads to charges of "double dipping"
 - tracking license through metadata is hard
- Diamond
 - Content goes through the normal publishing process, but
 - fees are supported though public or charitable funds,
 - authors experience no need to arrange payment
 - large question over long term sustainability of this model

Creative common licenses

- CC0
- CC-BY
- CC-BY-NC
- CC-BY-NC-ND

Funders and Funder mandates

- (2013) Wellcome Trust
- (2013) White House Memo
 - (1991) ArXiV
 - (2000) PubMedCommons
 - (2013) BioArXiV
 - (2014) CHORUS Launched 2014
- (2014) HEFCE
- (2014) Bill & Melinda Gates Policy
- (2015) NRO

2014 over 500 OA policies, growing at
16% y/y

publishers

- BMC
- PLOS
- eLife
- PeerJ
- University Presses

publishers

- BMC
- PLOS
- eLife
- PeerJ
- University Presses
- Elsevier
- SpringerNature

has OA been a success?

YES

- more content is available free to read
 - from 2003 to 2013 OA PubMed content grew from 1% to 15%
- most funders have accepted that OA is in line with their mission
- many new journal launched today are OA journals

if so, why?

- commerical publishers have discovered that it can be profitable
- it appeals to our better angels
- it strongly aligns with ideas of the public good
- the idea of creating a well functioning marketplace is also an appealing one

NO

- global costs in the industry have continued to rise, and not been checked
- commercial publishers have retained an almost monopoly position in the market
- most authors today remain uninterested in OA for its own sake
- it's only captured about 2% of the revenue in the market

if not, why not?

- there remains a collective action problem
- publishing infrastructure is baroque (including XML workflows)
 -
- direct to web solutions have not disrupted the existing scholarly publishing infrastructure
- journals remain a collection of micro-monopolies

cost of transition to a more efficient system is high

meaning

open access

Oxford xml summer school 2015

OA publishing is a service based model, recouping the costs can be done via:

- APCs the main customer is the author/funder
 - membership model
 - direct support model (diamond)
 - making the cost of production disappear

considered a cost of the research
process

OA publishing is a service based model, recouping the costs can be done via:

- APCs the main customer is the author/funder
 - membership model
 - direct support model (diamond)
 - making the cost of production disappear

Subscription publishing is a content based model

- access: the main customer is the institutional librarian
- sponsorship
- advertising
- reprints

considered a cost of the research process

As long as you are creating more value
then you are capturing, then you are
doing good for your community.

What services does the customer want?

- the main customer is the author

author needs can still be served by XML workflows, but the rest of the stack is not serving them well.

they want:

- fair and balanced peer review
- rapid publication times
- a “good publication”
- a discoverable publication
 - indexing in google scholar
 - coverage in scopus/web of science
 - discipline specific archiving
 - software that is respectful of their time

Research Data

research data

Oxford xml summer school 2015

Why care about data?

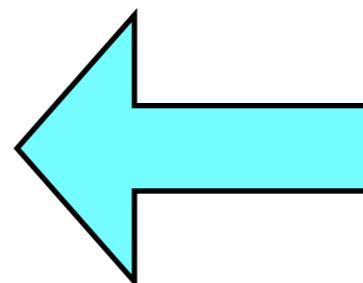
It's obvious to say this, but the data that research is built on is an integral part of the scientific process.

What's the relationship between the paper, and data?

A fair criticism of current publishing workflows is that publishers force authors to reduce much of their raw data into the form of graphics or images, and then to compress these images into low resolution versions for inclusion in a manuscript. The reader wants to get the underlying data, but has to go through a lot of hoops to get it.

Patterns

- don't connect the data to the paper at all
- refer obliquely to to the data set in the body of the publication
- link to the data set in the body of the publication via uri /identifier
- dump the data into supp info
- deposit in data cite, and hope there is a link to the paper
- link to the paper from the dataset
- create a specific section of the paper tagged about data
- cite the data in the reference list
- enhance the metadata of the paper in crossmark pointing to the data
- create a micro publication
- create a meta-paper about the data



Ugh



Meh



Yeah!

- It's one component of a research object!
 - Article
 - Data
 - Code
- It's one component of a story from the author
 - Preprints
 - Posters
- It's an evolution of a narrative
- versioning and changing research objects
 - giving a DOI to each revision of an article?
 - giving a DOI to each item of a dataset?
- The relationship between different components can be signalled.

Scales of data

- BIG DATA
- small data
- data that is starting to get out of control

Flavours of data

- Well structured
- semi-structured
- totally unstructured

How to handle data?

- We don't really know
- some data journals coming to market
- now exists a way to cite data, but not widely used yet
- Funders increasingly looking to bring in open data policies

Where can you put the data, and what can you find out?

Data repo	Metrics	API
EBI/PDG/BGI ...	?	Many
Figshare/projects	views/downloads/shared	REST Oauth
Imeji	web metrics	RDF
Dryad	views/downloads	OAI-ORE/PMH RDF
Datacite affiliated store		REST Basic HTTP Auth, OAI
Zenodo	Altmetric	OAI-ORE/PMH
Lab archives	No	Yes
Dataverse	web metrics	REST Basic HTTP Auth, OAI
Github	pull requests forks following	REST Oauth
Amazon	usage cost	REST keys based
Lab cluster	No	No

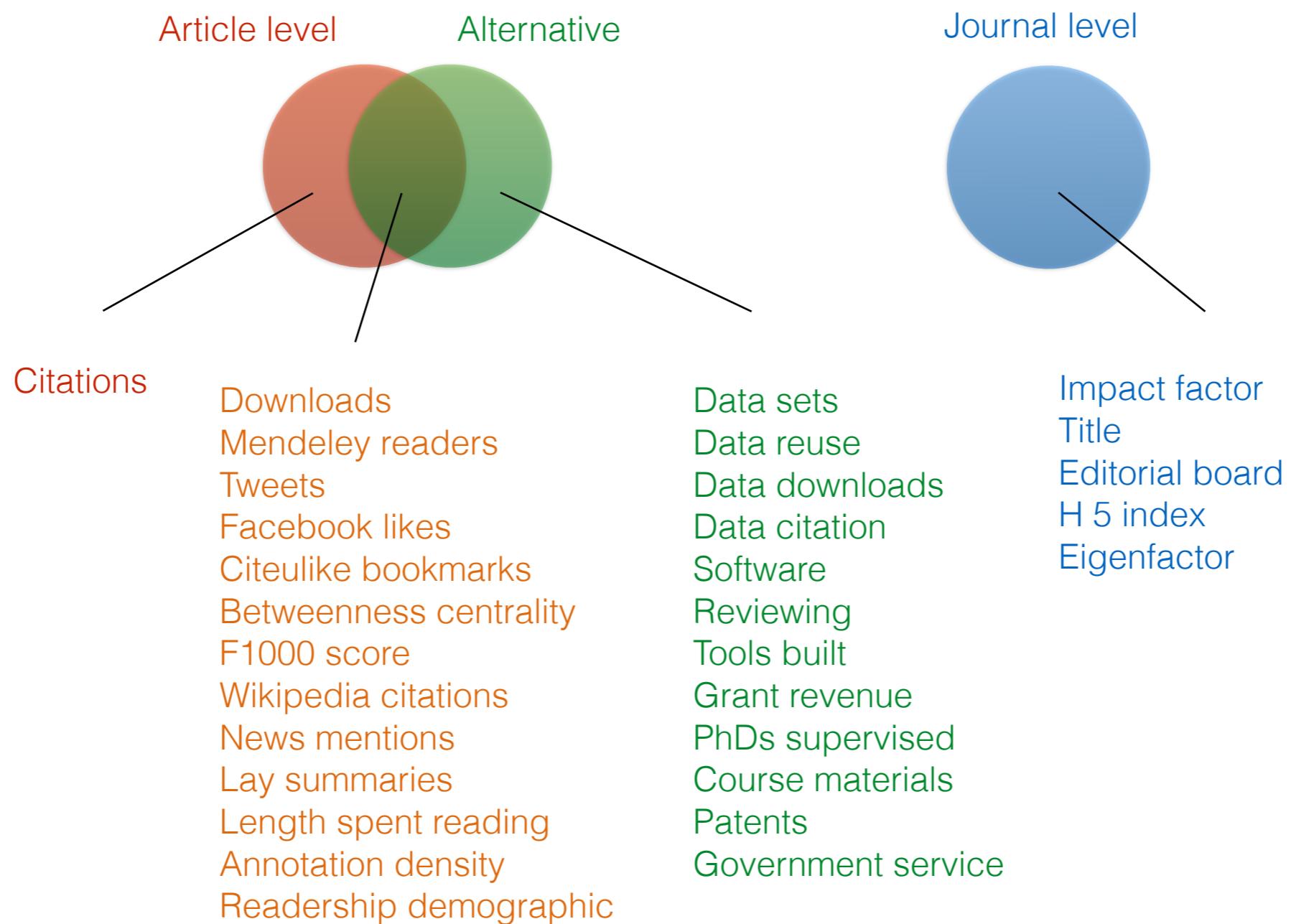
Getting credit for data

AltMetrics

altmetrics

Oxford xml summer school 2015

what are they?



what's their history?

Manifesto was written in 2005

Bibliometrics have been around for a lot longer

The impact factor might be the first citation metric to be developed (1950s),
but it was restricted to journal counts due to technical limitations.

what's the current state of the art?

- altmetric.com
 - Nature
 - Springer
 - Wiley
- lagotto - the open source tool
 - PLOS
 - copernicus
 - eLife
- plumb analytics
- publishers creating or running their own platforms
 - frontiers
- repositories adding UIDs
 - figshare
 - datadryad
 - zenodo

what's next for ALMs?

- They provide a variety of different kinds of value to different actors in the space, but there has not yet been a “killer app”
- They allow researchers to create narratives of their work and impact (impactstory).

but what are they really about?

- They are really about problems in the assessment and rewards system in academia.
 - see DORA - the declaration on research assessment

What have they got to do with XML, markup and structure?

- data citation
- well identified objects help
- objects that show their interrelationship help
- the fuller a digital life that an object has, the easier it is to automate the gathering of information about that object, that can then facilitate the creation of stories around that object.

emerging trends

Reproducibility

Webkit and the browser

HTML5/Markdown/JSON

Oxford xml summer school 2015

Reproducibility

reproducibility

Oxford xml summer school 2015

What's the issue with reproducibility?

AMGEN result - 88% could not be replicated

2011 - 80% of studies could not be replicated

nature editorial

Efforts to address this problem

- computable papers
- retraction watch
- centre for open science - Reproducibility project

Reproducibility Project Cancer Biology

- RPCB
 - our process
 - paper selection
 - registered reports
 - peer reviewed replication studies

criticisms

- expense
- negative connotations of failing to get your paper replicated
- waste of effort, more effort should be put into novel research

why can't papers get replicated?

- lack of sharing of data
- poorly described protocols
- the experiment is just really hard

The role of the publisher

- push for open peer review
- make all data open and available
- implement data checking in the review process (e.g. EMBO checks for image manipulation)
- experiment with new ways of presenting papers and data

Article Figures & Data Metrics Article & Author Info

Cdc48/p97 promotes degradation of aberrant nascent polypeptides bound to the ribosome

Rati Verma , Robert S Oania, Natalie J Kolawa, Raymond J Deshaies 

California Institute of Technology, United States; Howard Hughes Medical Institute, California Institute of Technology, United States

DOI: <http://dx.doi.org/10.7554/eLife.00308>

I-1 Abstract

Ubiquitin-dependent proteolysis can initiate at ribosomes for myriad reasons including misfolding of a nascent chain or stalling of the ribosome during translation of mRNA. Clearance of a stalled complex is required to recycle the ribosome for future use. Here we show that the ubiquitin (Ub) pathway segregase Cdc48/p97 and its adaptors Ufd1-Npl4 participate in ribosome-associated degradation (RAD) by mediating the clearance of ubiquitinated, tRNA-linked nascent peptides from ribosomes. Through characterization of both endogenously-generated and heterologous model substrates for the RAD pathway, we conclude that budding yeast Cdc48 functions downstream of the Ub ligases Ltn1 and Ubr1 to release nascent proteins from the ribosome so that they can be degraded by the proteasome. Defective RAD could contribute to the pathophysiology of human diseases caused by mutations in p97.

DOI: <http://dx.doi.org/10.7554/eLife.00308.001>

I+1 eLife digest

   
Aa Aa Aa

[DOWNLOAD CITATIONS](#)

[SHARE](#)

Jump to:

- Article
 - Abstract
 - eLife digest
 - Main text
 - Introduction
 - Results
 - Discussion
 - Materials and methods
 - Decision letter (highlighted)
 - Author response

[Figures & data](#)

[Metrics](#)

[Article & Author Info](#)

Categories & tags

[Research article](#) [Biochemistry](#)

[Cell biology](#) [ubiquitin](#) [ribosome](#) [Cdc48](#)

I-1 Decision letter

Ivan Dikic, Reviewing editor, Goethe University, Germany

eLife posts
the approv
indicatin
opportunity
typical sh

Thank you for
nascent poly
evaluated by
decision has
to reveal his
General ass
targeting na
complex pla
Cdc48 has t
substrate co
because the
pathways. A
mechanistic
that the Cdc
aspects of th
and a better

DOI: <http://dx.doi.org/10.7554/eLife.00000>

I-1 Author response

The main changes from the original submission are as follows:

1. The main criticism of the original submission was that there was not sufficient mechanistic insight into how Cdc48–Ufd1–Npl4 promotes degradation of tUNCs and non-stop decay (NSD) pathway substrates that are translated from messages that lack a stop codon. In particular, we provided no evidence to indicate that the role of Cdc48–Ufd1–Npl4 in this process is direct. In a series of email and telephone discussions with the editors, we were advised that providing evidence to support a direct role for Cdc48–Ufd1–Npl4 in RAD might suffice to address the concern that the reviewers raised about mechanism. We have addressed this criticism by showing in the revised manuscript that Cdc48 and Ufd1 were associated specifically with stringently-washed, affinity-purified ribosomes (Figure 1D). We also show that Cdc48 and Ufd1 bound the NSD substrate GFH^{NS}, but not the control GFH^{Stop} (Figure 3E), despite the latter being present at much higher levels than the NS reporter.
2. In addition to the new figure panels described above (Figures 1D and 3E), we have added the following additional data panels to address the other criticisms made by the reviewers: (i) input controls for Figure 1B, (ii) evidence that accumulation of tUNCs in *cdc48-3* cells is reversed by expression of wild type Cdc48 but not the ATPase-deficient Q2 mutant (Figure 1C), (iii) an expanded Figure 2B to include a control in which RNase was added to cell lysate prior to isolation of ribosomes, (iv) a new version of Figure 3A showing the effect of *ubr1Δ* and *ltn1Δ* on accumulation of tUNCs in *cdc48-3* mutants, (v) a new version of Figure 3B showing the effect of Cdc48 pathway mutations on accumulation of the NSD substrate GFH^{NS}, including an anti-tubulin loading control, (vi) degradation assays for GFH^{NS} in wild type, *ltn1Δ*, and *ufd1-2* cells and GFH^{Stop} in wild type cells (Figure 3C), (vii) high-resolution sucrose gradient fractionation of lysates of *ltn1Δ* and *cdc48-3* cells expressing PrA^{NS} (Figure 4B; these data replace the sucrose gradient

Fractionations of tUNCs that were previously shown in Figure 2D. Total levels of Cdc48-PTC in total cell

The browser

the browser

Oxford xml summer school 2015

The browser

- Safari
- Electron

- Atom
- Slack
- Pixate Studio - demo

Classic Steps In Publishing

Most of these steps can now be implemented in the browser, or in a browser based app.

- Authoring
- Submission
- Peer review
- Document conversion
- Content enhancement
- Annotation
- Corrections cycle
- Pdf / print / pagination
- Downstream applications
- Search

- Authoring
 - Manuscripts.app - DEMO
 - substance writer - DEMO
 - gitbook - DEMO
- Submission
 - Overleaf extension
- Peer review
- Document conversion
 - see lens demo in next session
- Content enhancement
- Annotation
 - hypothesis
- Corrections cycle
 - screenshot of tool from Ravi
- Pdf / print / pagination
 - vivlio style demo
- Downstream applications
 - HTTP / FTP
- Search
 - elasticsearch demo

Markup Markdown/ HTML5

markup/markdown/html5

Oxford xml summer school 2015

This will go into more detail on some document conversion steps that might be interesting to this group.

- <http://scholdoc.scholarlymarkdown.com>
<http://blog.martinfenner.org/2015/04/23/introducing-scholarly-markdown-bundle/>
- pandoc
- lens
- substance writer
- paged media standard

(vivliostyle demo) (<http://www.vivliostyle.com/en/sample/>)

markup/markdown

Oxford xml summer school 2015

Lens Demo

markup/markdown/html5

Oxford xml summer school 2015

Slides Demo

markup/markdown/html5

Oxford xml summer school 2015

Speculative Trends

Computational Papers

Living Papers

The Blockchain

Oxford xml summer school 2015

Computational papers

- encode project
- iPython/Project Jupyter
- nbviewer
- Bio.JS
- FigShare

Living Papers

- wikipedia
- eLife research advances
- <http://theoj.org>
- ReScience
- f1000 research

Advances demo

living papers

Oxford xml summer school 2015

Article

Figures & data

Metrics

Article & author info

Beam-induced motion correction for sub-megadalton cryo-EM particles

Sjors HW Scheres 

Medical Research Council Laboratory of Molecular Biology, United Kingdom

DOI: <http://dx.doi.org/10.7554/eLife.03665>

Published August 13, 2014

Cite as eLife 2014;3:e03665

I-1 Abstract

In electron cryo-microscopy (cryo-EM), the electron beam that is used for imaging also causes the sample to move. This motion blurs the images and limits the resolution attainable by single-particle analysis. In a previous Research article (Bai et al., 2013) we showed that correcting for this motion by processing movies from fast direct-electron detectors allowed structure determination to near-atomic resolution from 35,000 ribosome particles. In this Research advance article, we show that an improved movie processing algorithm is applicable to a much wider range of specimens. The new algorithm estimates straight movement tracks by considering multiple particles that are close to each other in the field of view, and models the fall-off of high-resolution information content by radiation damage in a dose-dependent manner. Application of the new algorithm to four data sets illustrates its potential for significantly improving cryo-EM structures, even for particles that are smaller than 200 kDa.

DOI: <http://dx.doi.org/10.7554/eLife.03665.001>[View Full Text](#)

I-1 Builds upon



Ribosome structures to near-atomic resolution from thirty thousand cryo-EM particles

Xiao-chen Bai, Israel S Fernandez, Greg McMullan, Sjors HW Scheres

A combination of direct-electron detectors and statistical movie processing allows ribosome cryo-EM structures to be determined to resolutions that were previously only attainable by X-ray crystallography.

eLife 2013;2:e00461

DOI: [10.7554/eLife.00461](http://dx.doi.org/10.7554/eLife.00461) View article with eLIFE Lens

Downloads:

 Article  Figures

Reference tools:

[DOWNLOAD](#) [OPEN](#)

Related content

Cryo-EM enters a new era

W Kühlbrandt

 — Biophysics and structural biology

Jump to:

Article

- [Abstract](#)
 - [Builds upon](#)
 - [Main text](#)
 - [Introduction](#)
 - [Approach](#)
 - [Results and discussion](#)
 - [Materials and methods](#)
 - [References](#)
 - [Decision letter](#)
 - [Author response](#)
 - [Leave a comment](#)
- [Figures & data](#)
- [Metrics](#)
- [Article & author info](#)



Article

Figures & data

Metrics

Article & author info

Ribosome structures to near-atomic resolution from thirty thousand cryo-EM particles

Xiao-chen Bai, Israel S Fernandez, Greg McMullan, Sjors HW Scheres 

Medical Research Council Laboratory of Molecular Biology, United Kingdom

DOI: <http://dx.doi.org/10.7554/eLife.00461>

Published February 19, 2013

Cite as eLife 2013;2:e00461

1–1 Abstract

Although electron cryo-microscopy (cryo-EM) single-particle analysis has become an important tool for structural biology of large and flexible macro-molecular assemblies, the technique has not yet reached its full potential. Besides fundamental limits imposed by radiation damage, poor detectors and beam-induced sample movement have been shown to degrade attainable resolutions. A new generation of direct electron detectors may ameliorate both effects. Apart from exhibiting improved signal-to-noise performance, these cameras are also fast enough to follow particle movements during electron irradiation. Here, we assess the potentials of this technology for cryo-EM structure determination. Using a newly developed statistical movie processing approach to compensate for beam-induced movement, we show that ribosome reconstructions with unprecedented resolutions may be calculated from almost two orders of magnitude fewer particles than used previously. Therefore, this methodology may expand the scope of high-resolution cryo-EM to a broad range of biological specimens.

DOI: <http://dx.doi.org/10.7554/eLife.00461.001>

[View Full Text](#)

1–1 Built upon by



Beam-induced motion correction for sub-megadalton cryo-EM particles

Sjors HW Scheres

Building on previous work (Bai et al., 2013), we describe an algorithm that allows cryo-EM structure determination to near-atomic resolution for protein complexes as small as 170 kDa.

eLife 2014;3:e03665

DOI: [10.7554/eLife.03665](http://dx.doi.org/10.7554/eLife.03665)

 View article with eLIFE Lens

Downloads:

 Article  Figures



Reference tools:

[DOWNLOAD](#) [OPEN](#)

Related content

[Direct detection pays off for electron cryo-microscopy](#)

N Grigorieff

 [Insight](#) — Biophysics and structural biology

Jump to:

Article

- [Abstract](#)
- [Built upon by](#)
- [eLIFE digest](#)
- [Main text](#)
- [Introduction](#)
- [Results](#)
- [Discussion](#)
- [Materials and methods](#)
- [References](#)
- [Decision letter](#)
- [Author response](#)
- [Comments \(1\)](#)

[Figures & data](#)

[Metrics](#)

[Article & author info](#)

ReScience demo

- <https://github.com/ReScience/ReScience/wiki>

The Blockchain & Science

the block chain and science

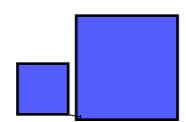
Oxford xml summer school 2015



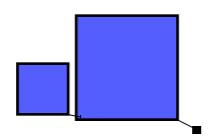
MB



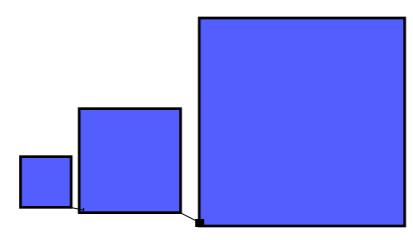
MB



MB GB



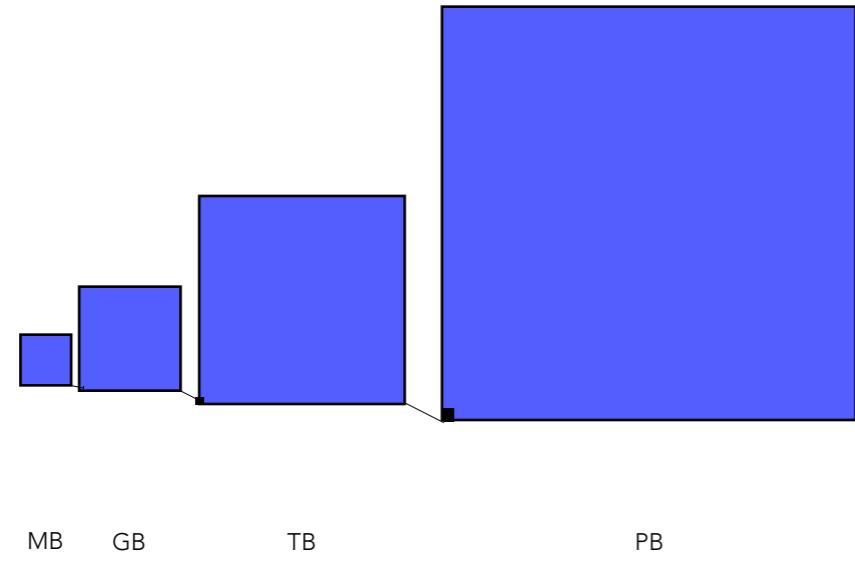
MB GB

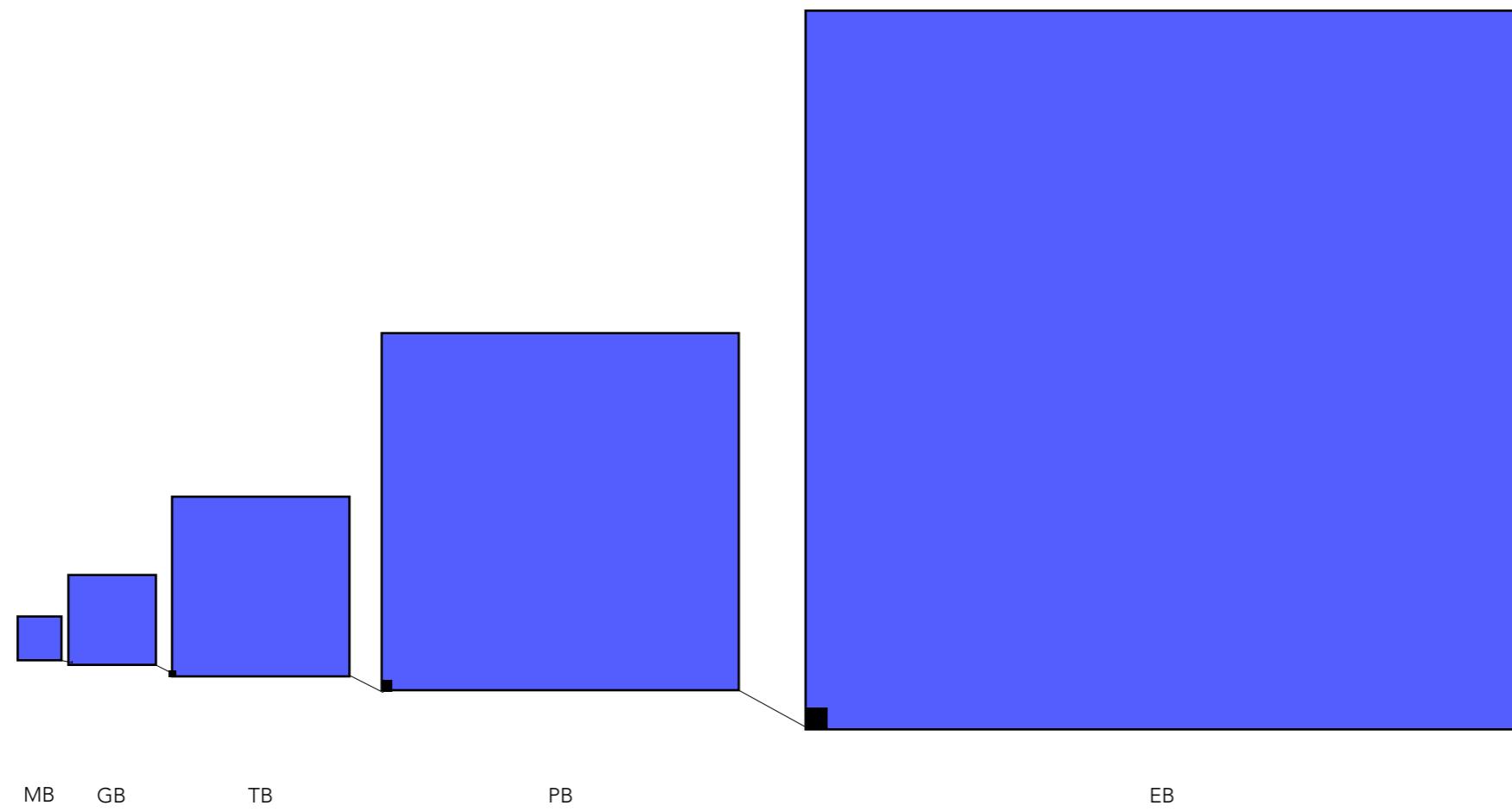


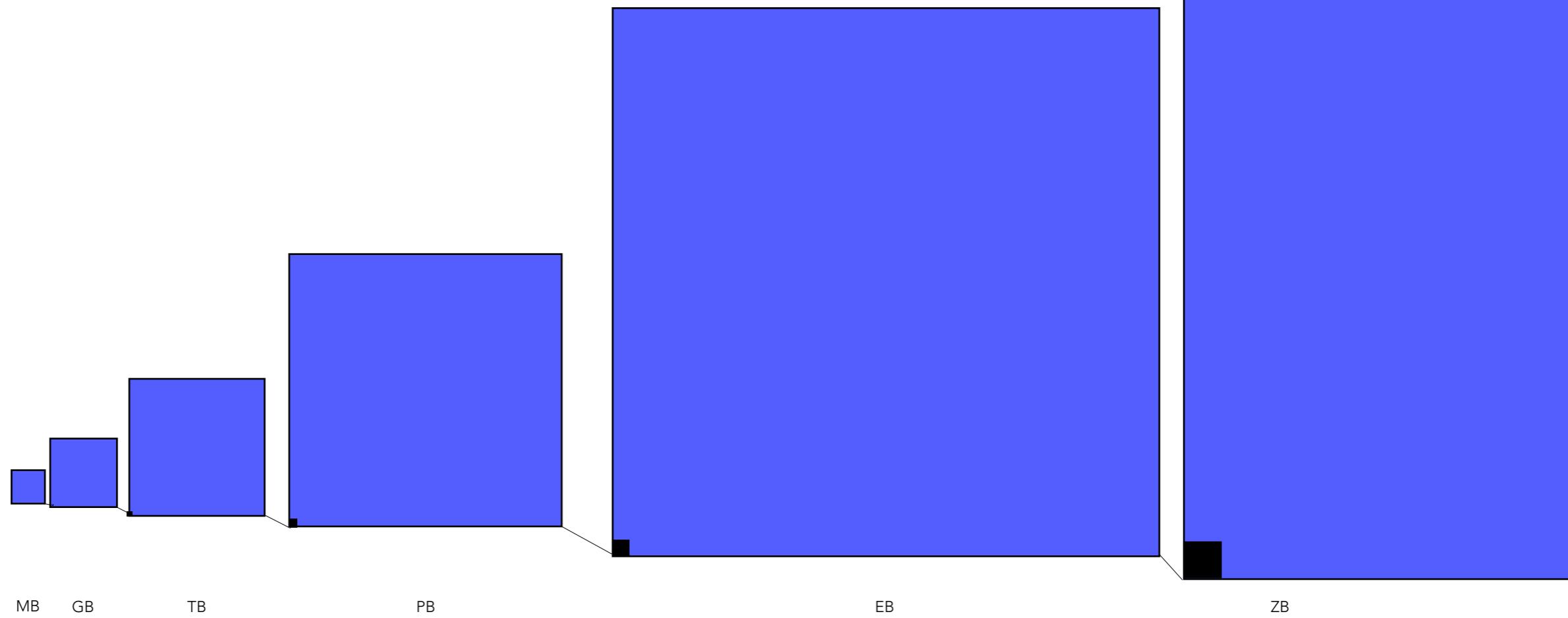
MB

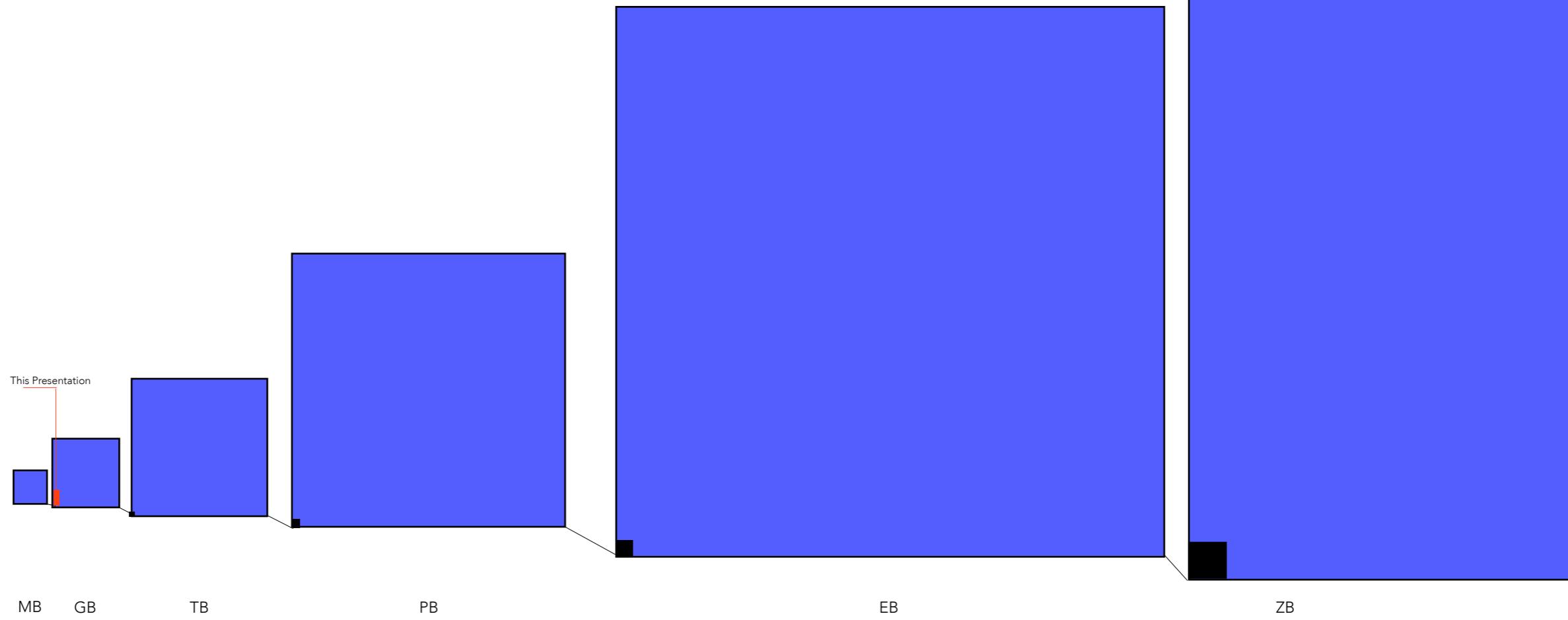
GB

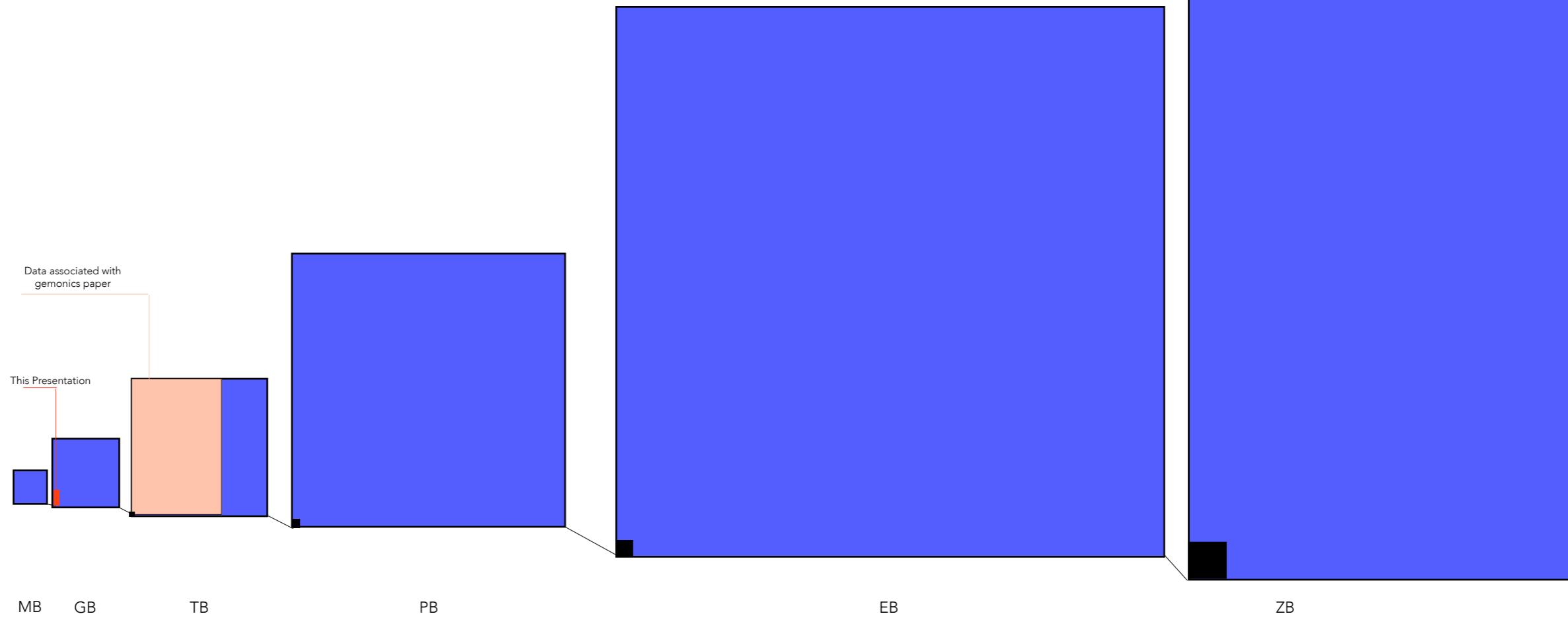
TB

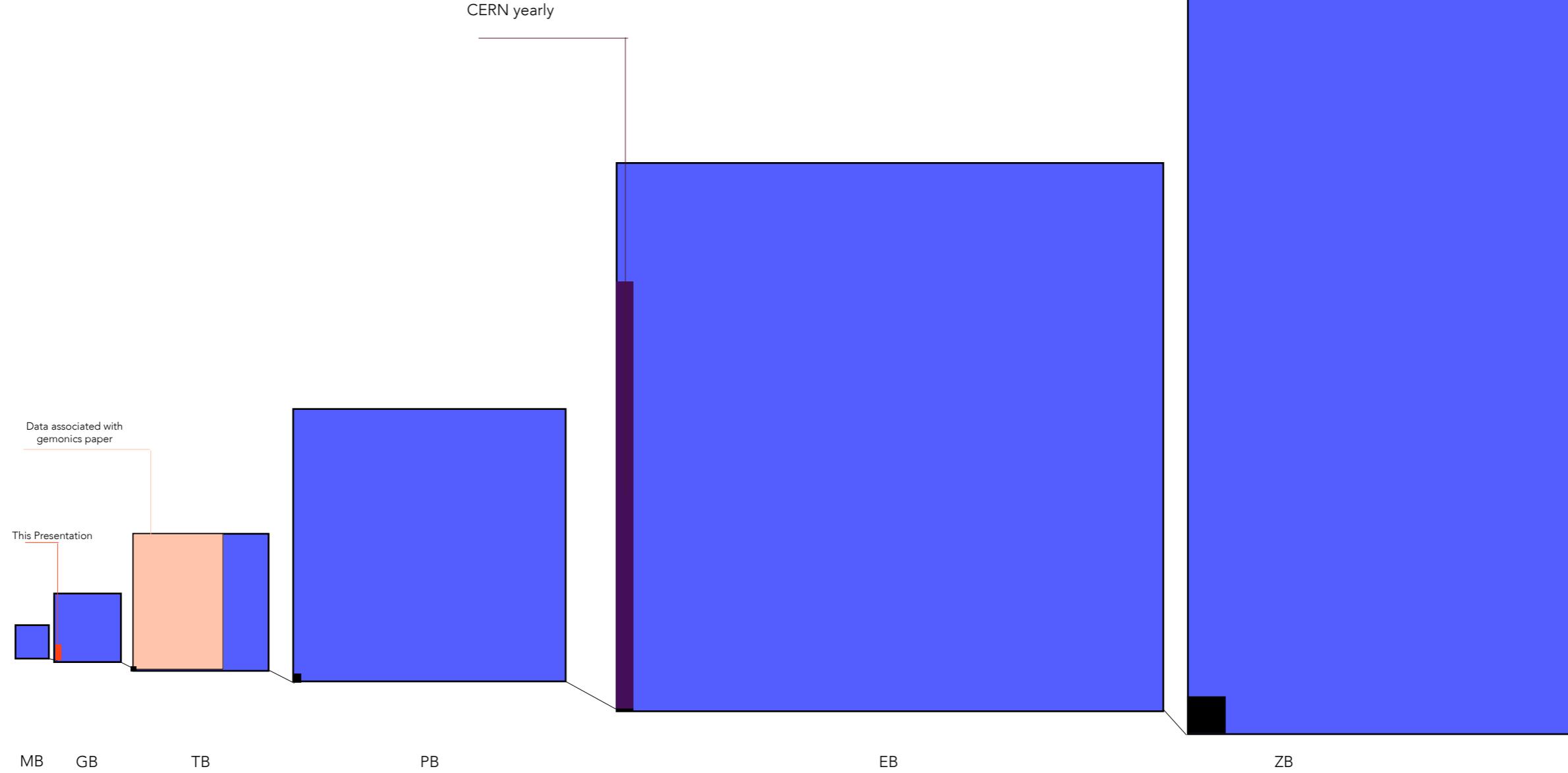


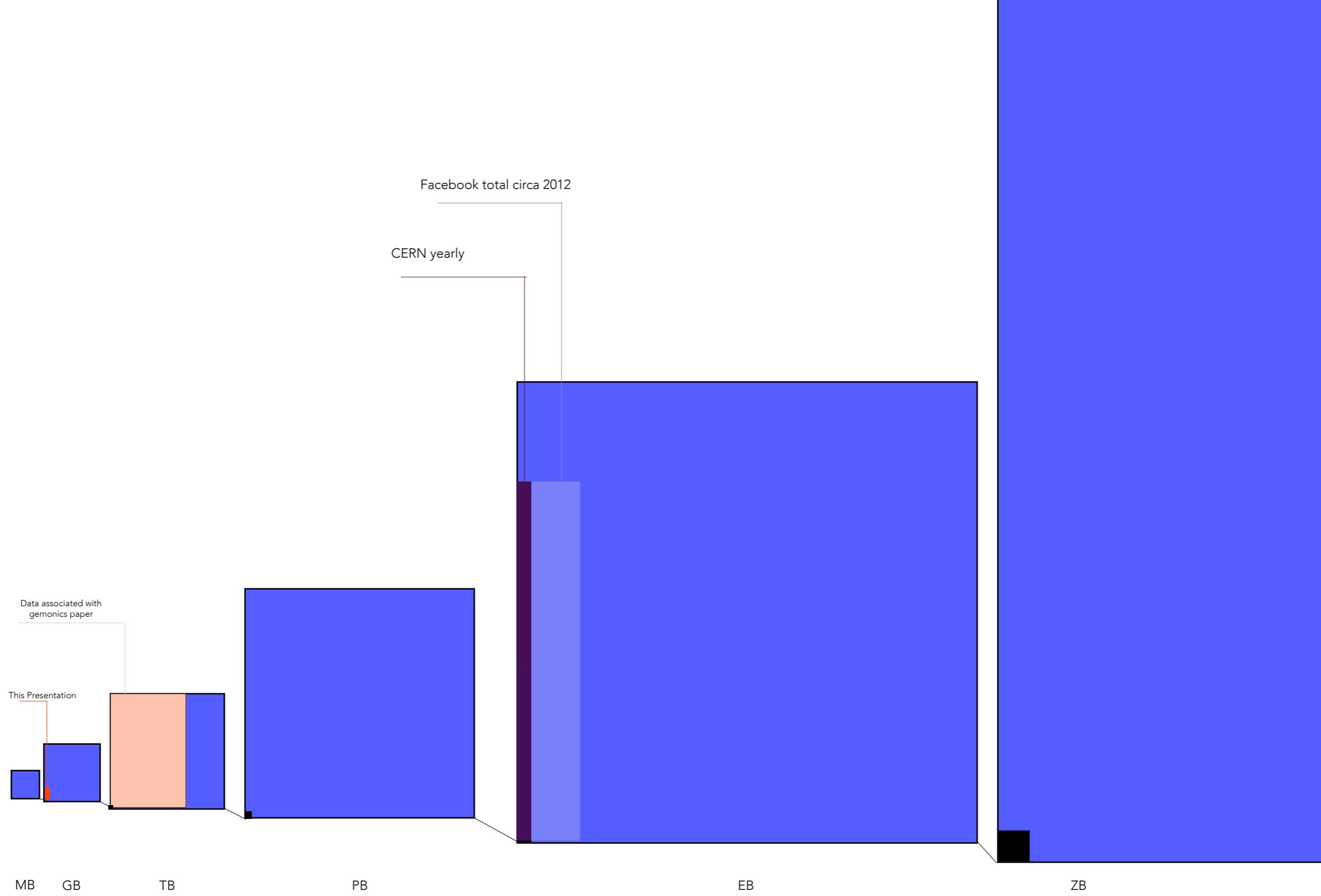




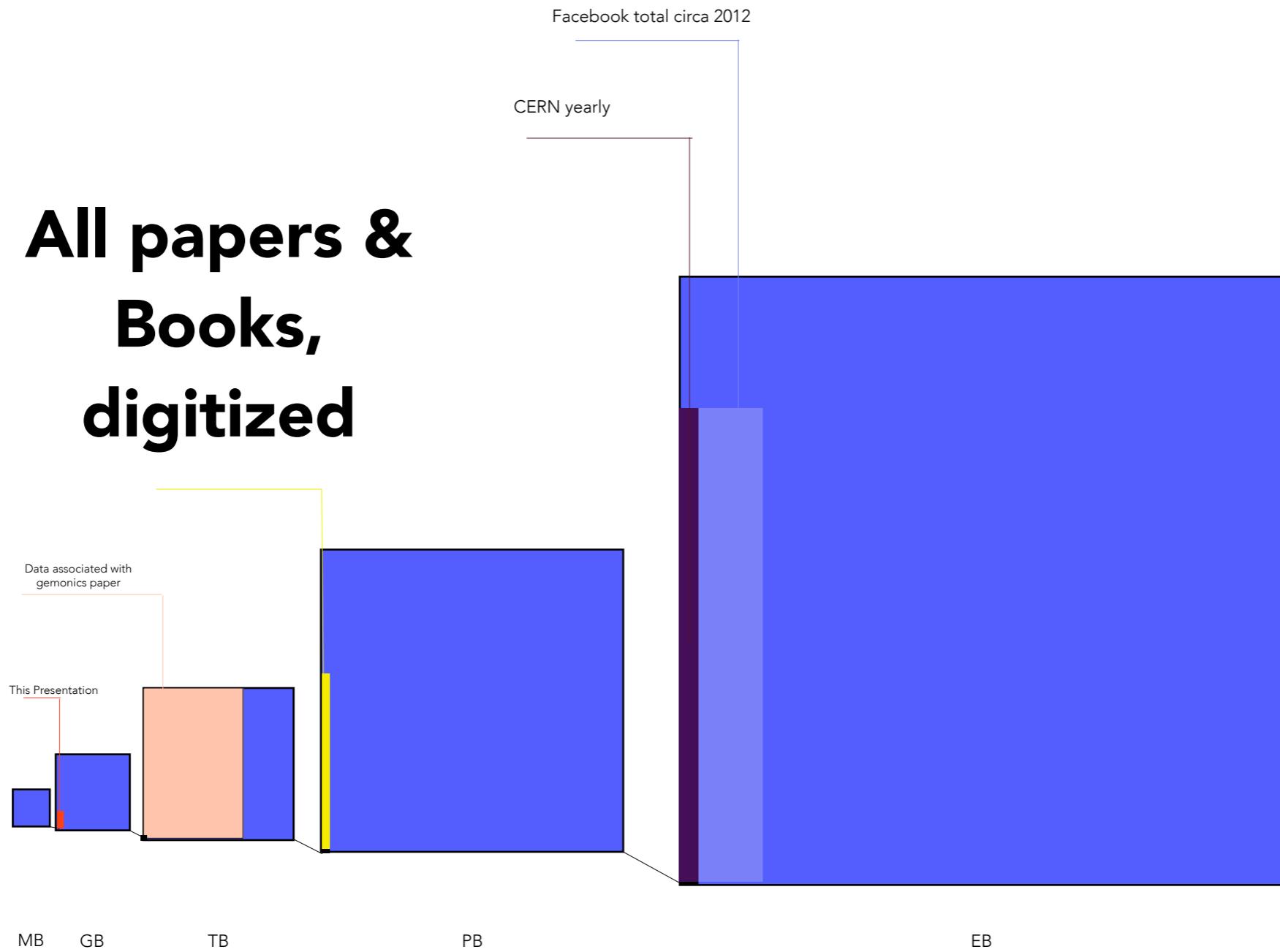




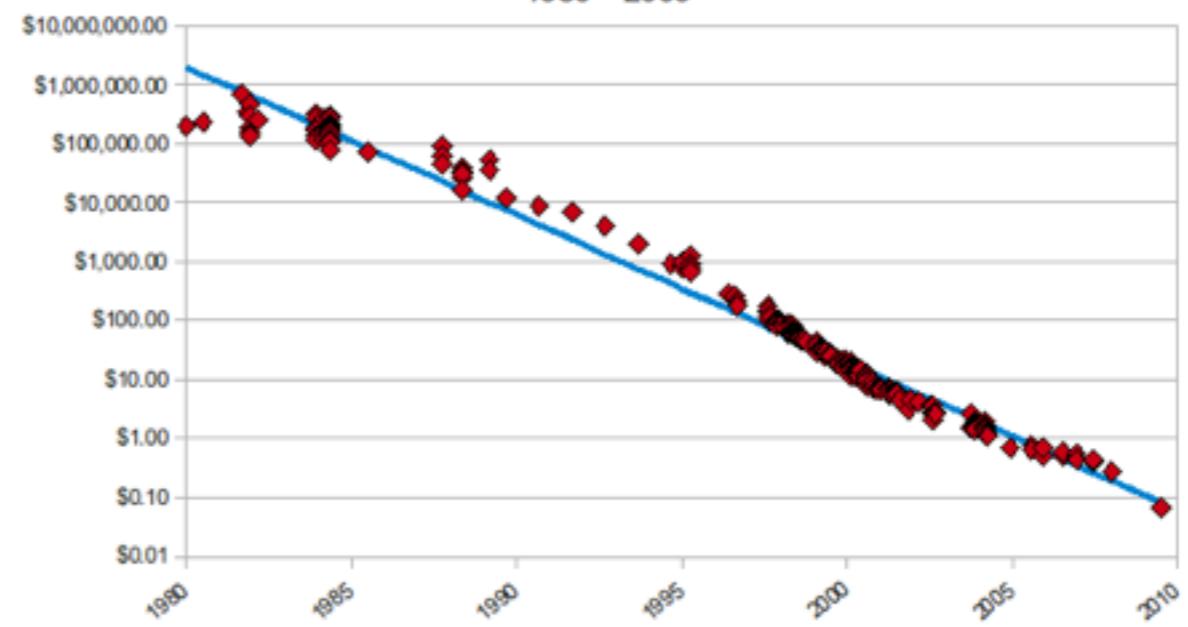




All papers & Books, digitized



Hard Drive Cost per Gigabyte 1980 - 2009



All papers & Books, digitized

