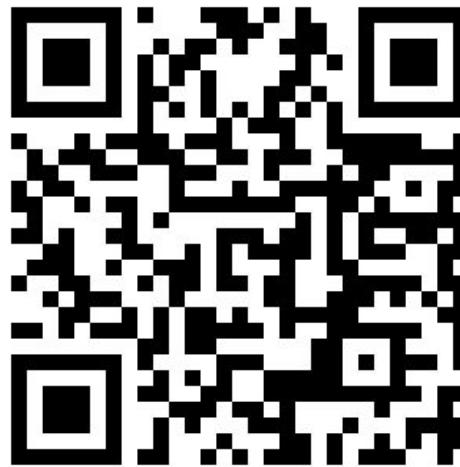


Chunked, Compressed, & Cloud-native
N-dimensional arrays

FOSS United x ILUG-D Meetup
18th May, 2024 @ IFF

Sanket Verma

Community and OSS @ Zarr



@MSanKeys963

Slides

https://bit.ly/zarr_foss_ilugd



GitHub Repository

https://bit.ly/foss_ilugd_more



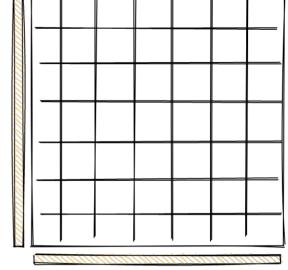
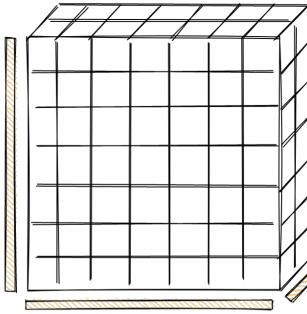
What I'll be talking about?

- What is Zarr?
- How is Zarr different?
- What is the Zarr Community and ZEPs?
- Demo

How Zarr works?

How Zarr works?

Arrays are container of items of the same data-type & size (in bits). The number of dimensions and items in container are described by the shape.

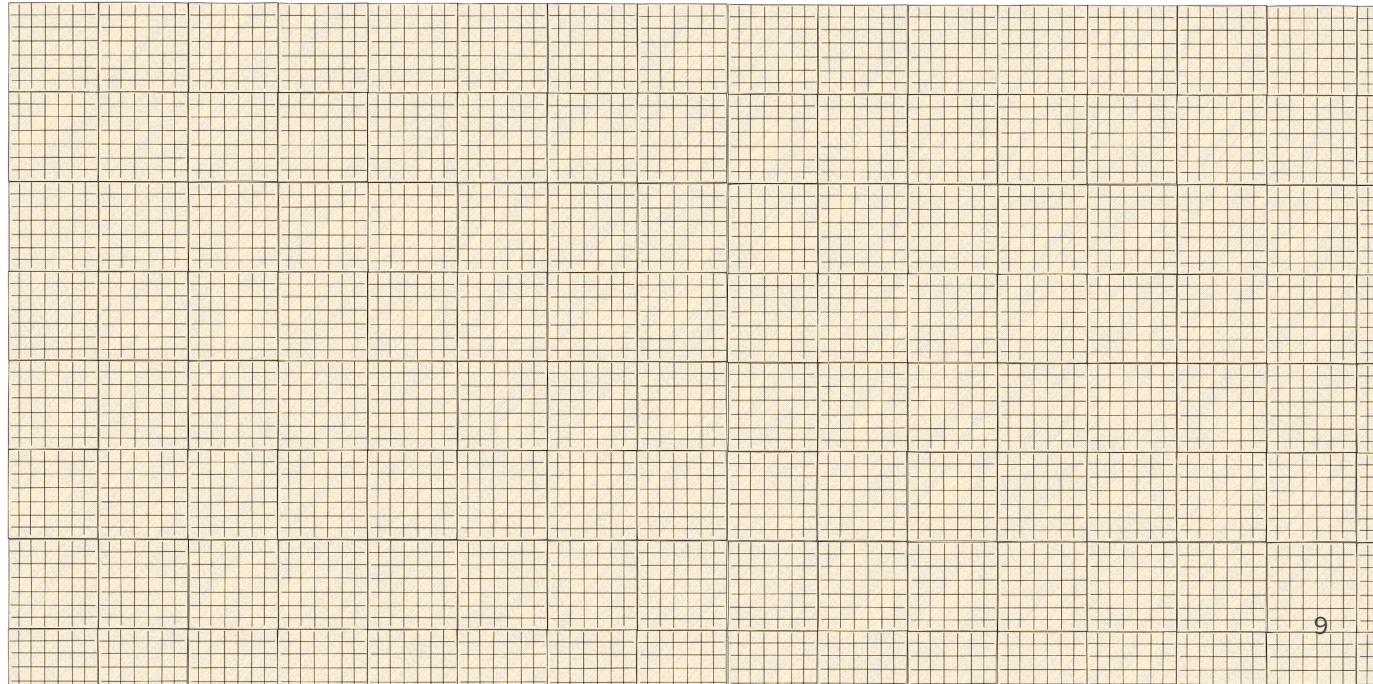
		
shape	(7,)	(7,7)
# dimensions	1D	2D
# items	7	$7 * 7$

How Zarr works?



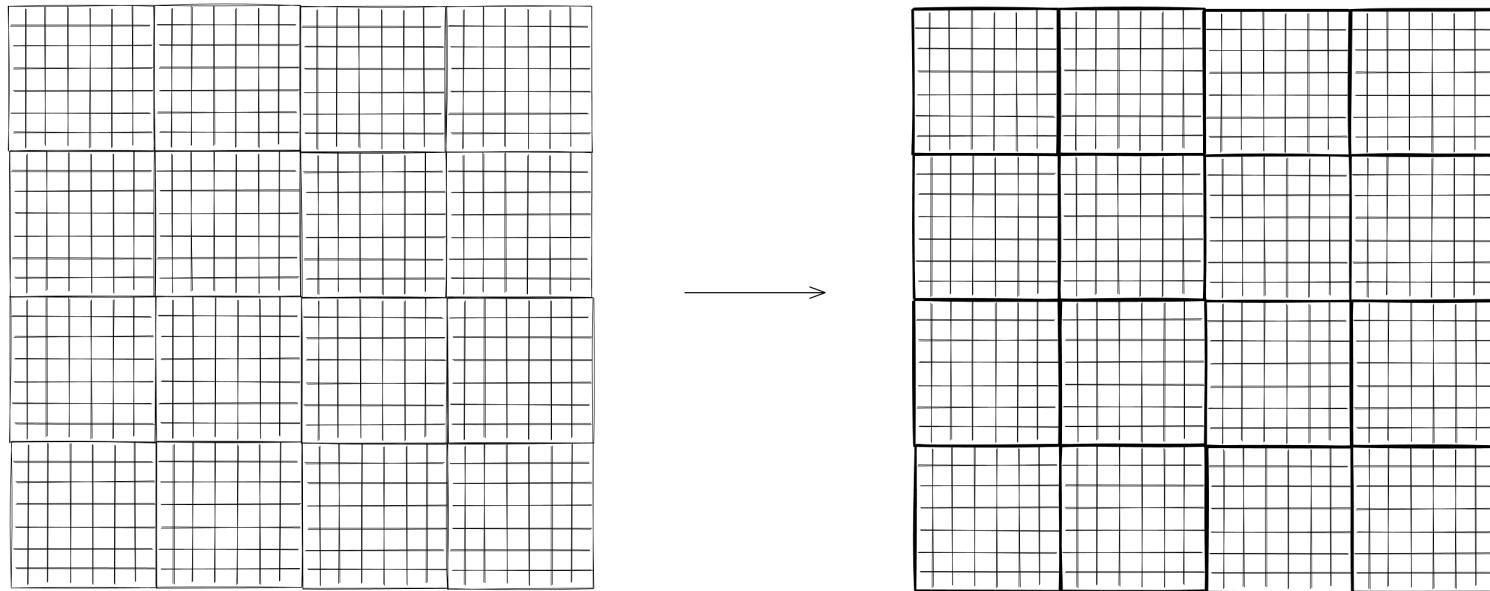
How Zarr works?

What if the data is too big to fit in memory?



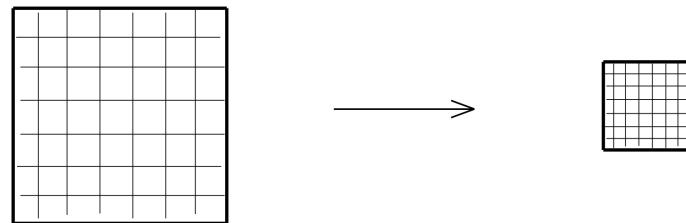
How Zarr works?

Divide array into chunks (Chunking)



How Zarr works?

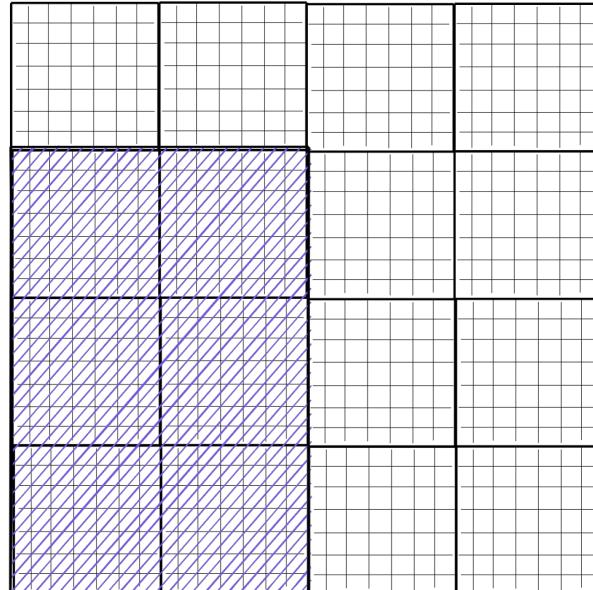
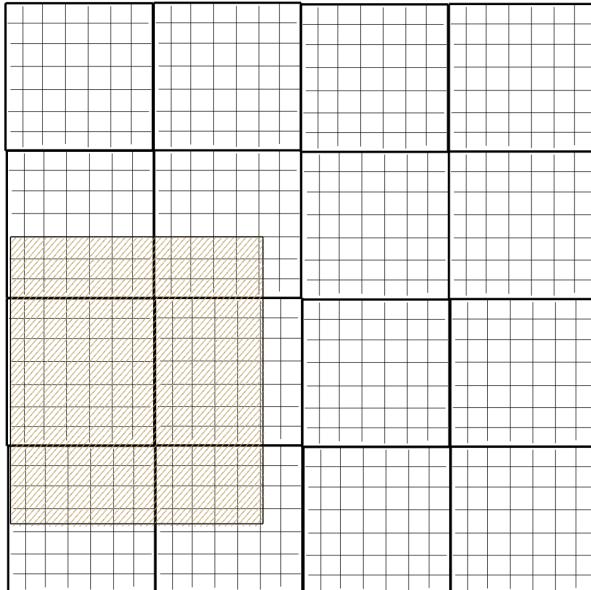
Compress each chunk



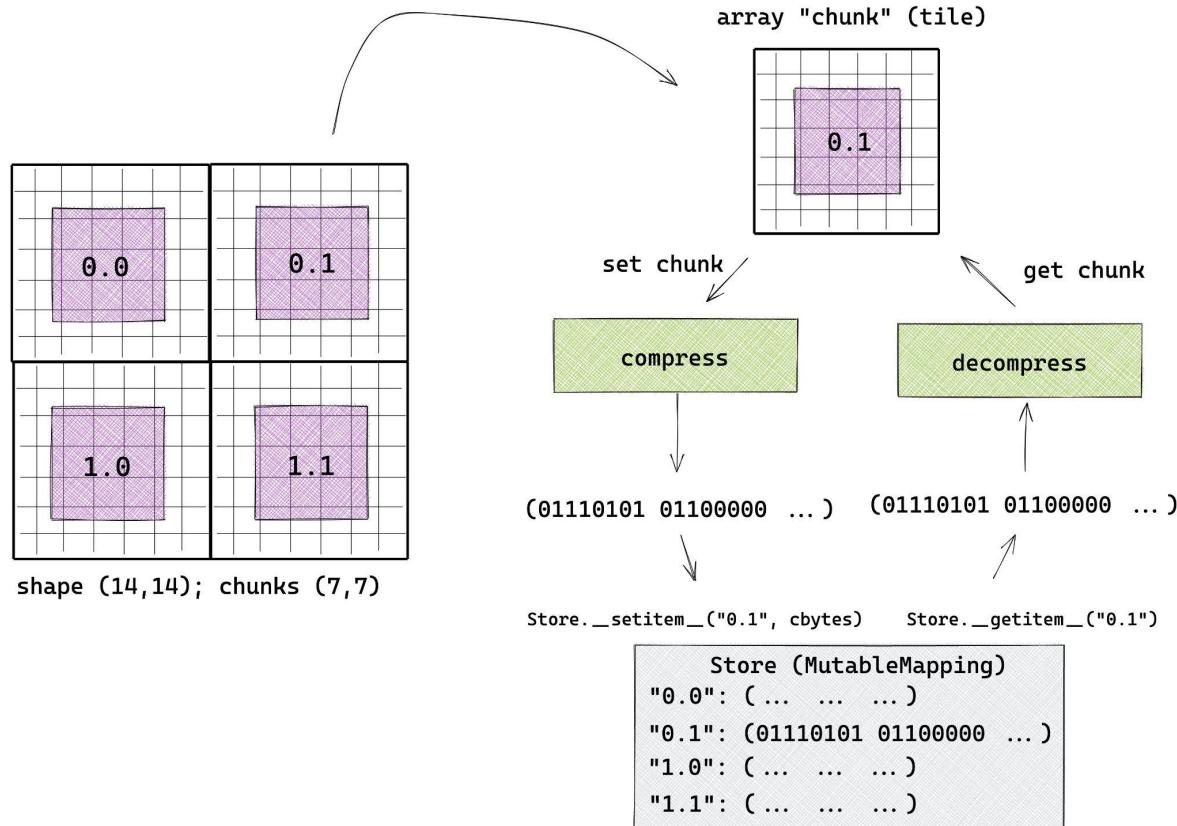
Over 20 supported compressors (BLOSC, Zstd, Zlib etc)

How Zarr works?

Retrieve chunks only when needed

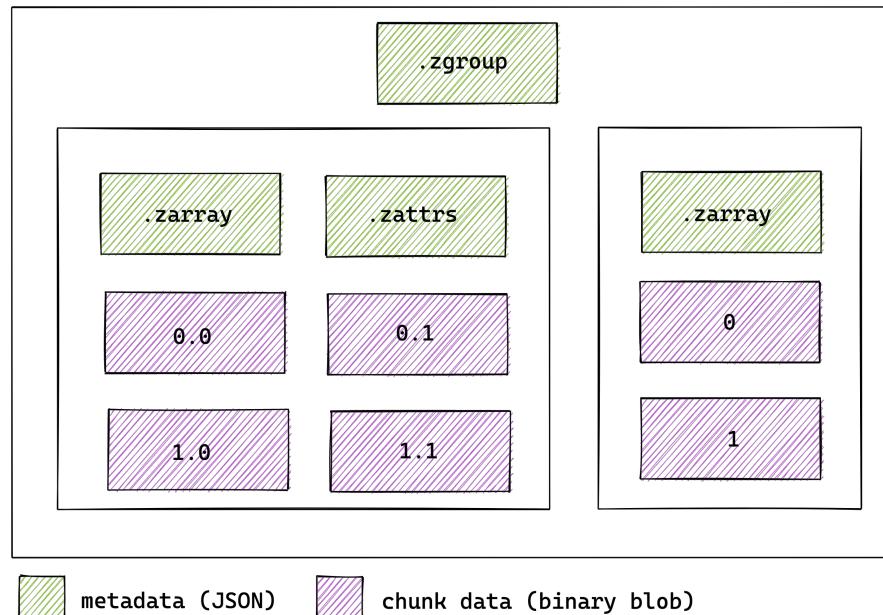


How Zarr works?



How Zarr works?

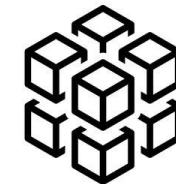
Multiple arrays can be organised in hierarchies of groups



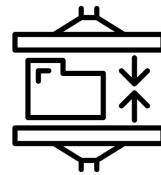
Benefits of Zarr



Distributed & Cloud Storage



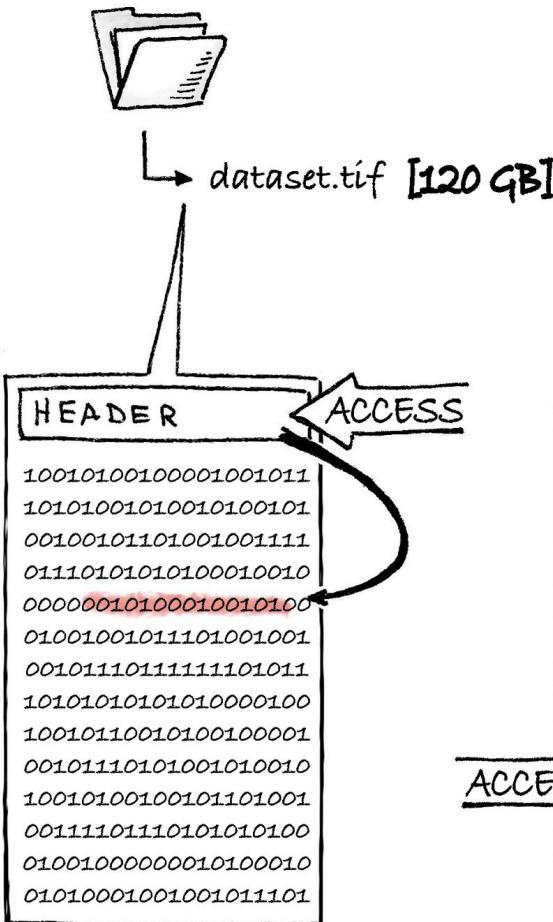
Chunked Storage



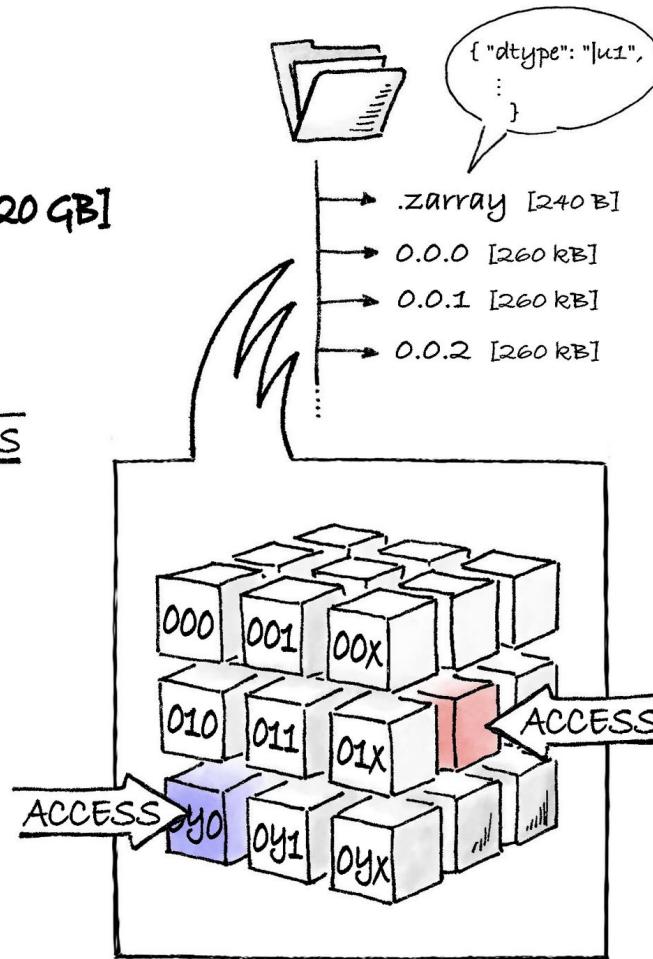
Built-In Compression

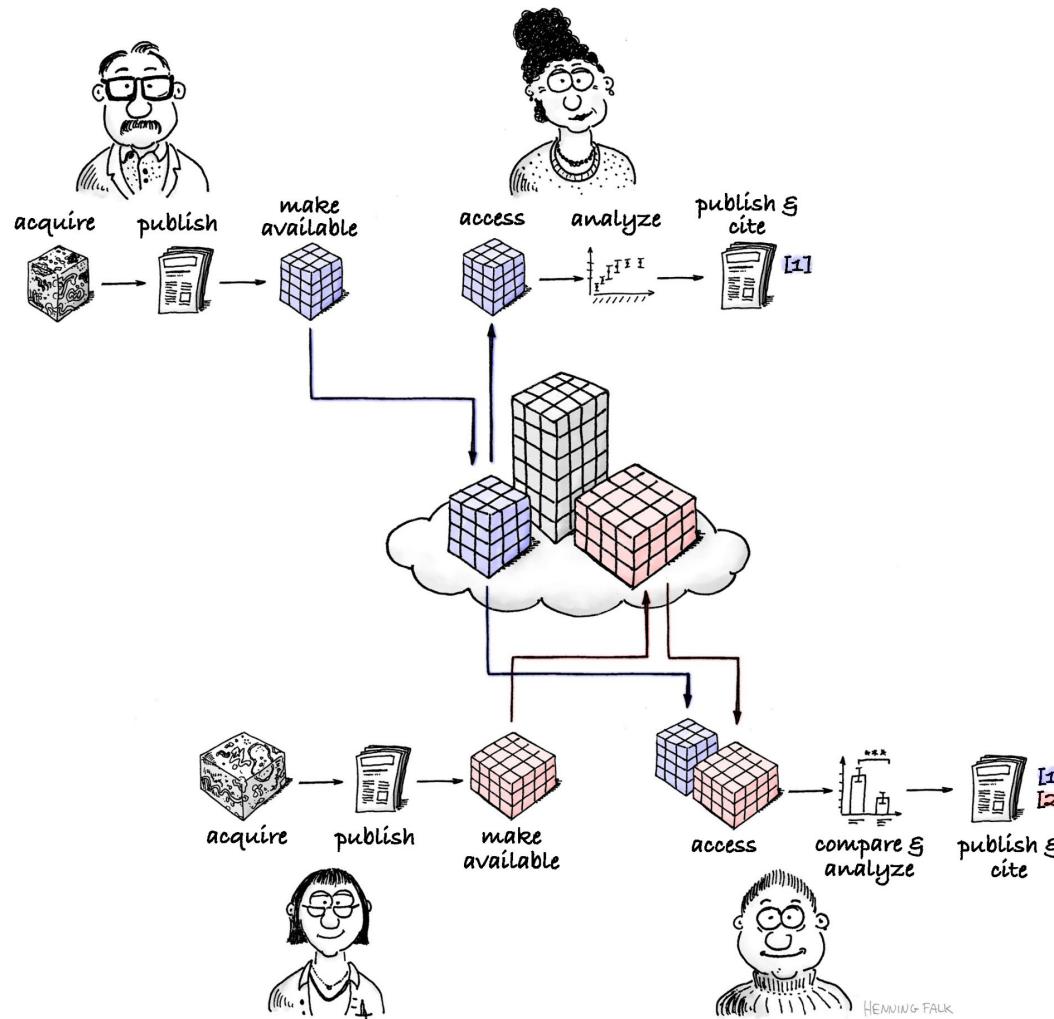
How Zarr is different?

Monolithic file format



OME-Zarr





When I was a lad, we
had to DOWNLOAD all
the data, before we could
look at it...!



More illustrations

Check here ➡



Zarr Specification

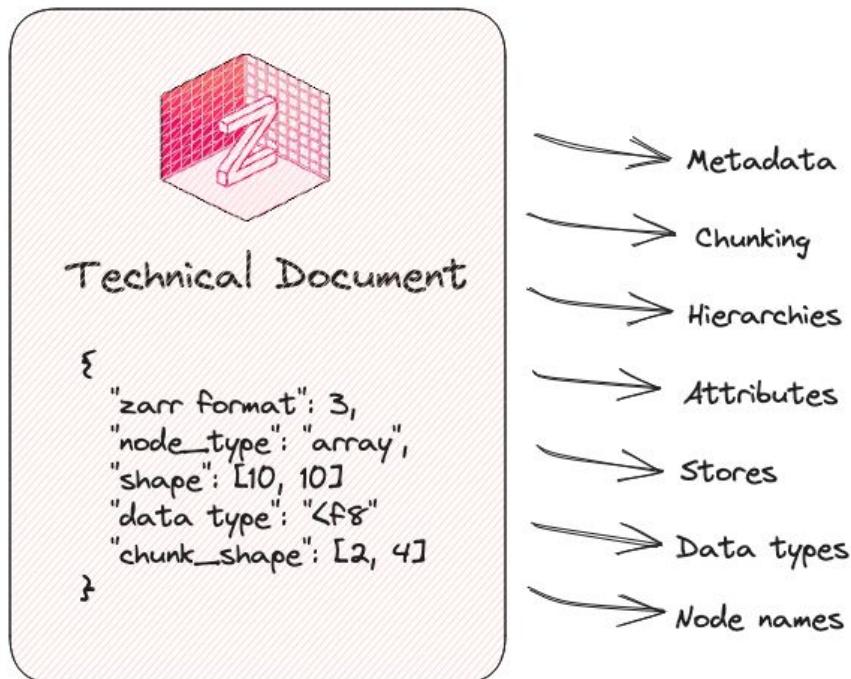
Zarr Specification

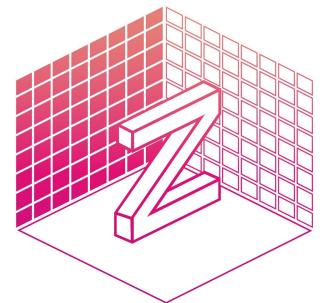
V1 → V2 → V3



<https://zarr-specs.readthedocs.io/>

Zarr Specification



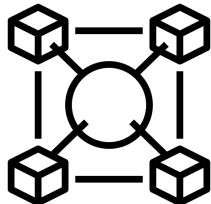


Zarr



Why did we work on V3 Specification?

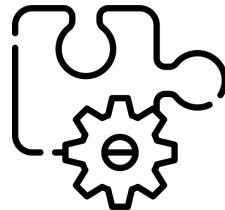
The Motivation



Interoperability



High-latency storage

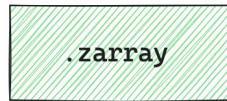


Extensibility

Major design updates

Restructured the
JSON metadata

Zarr Arrays → V2 vs. V3



0.0

0.1

0.2

0.3

1.0

1.1

1.2

1.3

2.0

2.1

2.2

2.3

3.0

3.1

3.2

3.3

V2



c/0

0

1

2

3

c/1

0

1

2

3

c/2

0

1

2

3

c/3

0

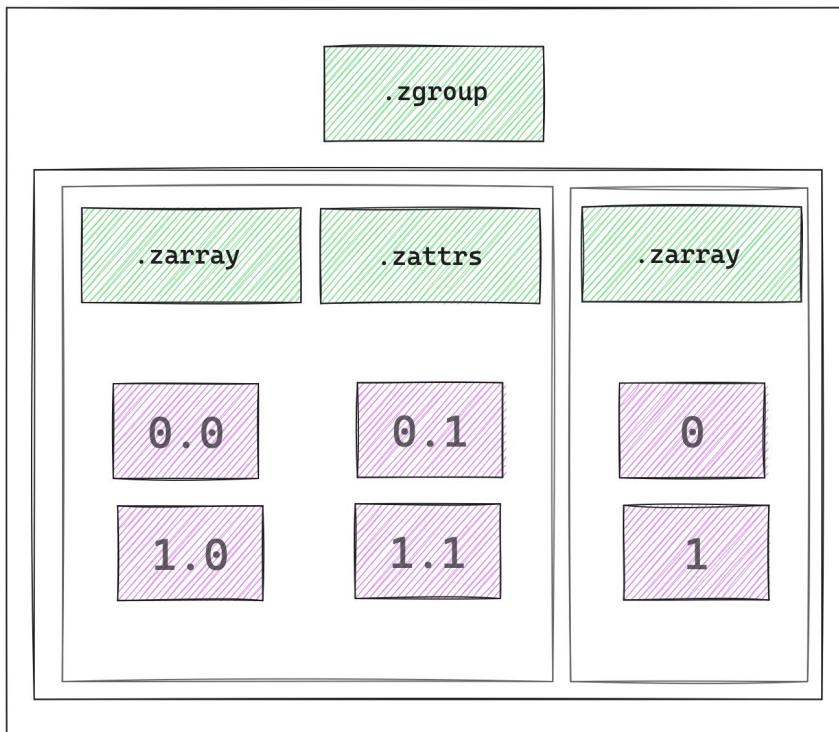
1

2

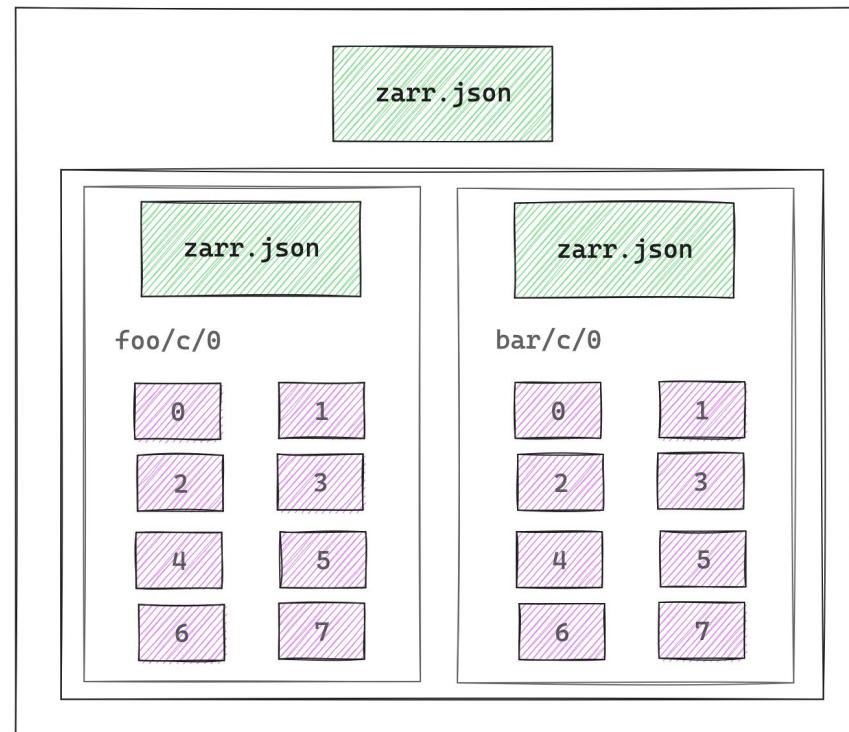
3

V3

Zarr Groups → V2 vs. V3



V2



V3

Zarr Community

ZARR OSS



AND SPECIFICATION

USERS & DEVS



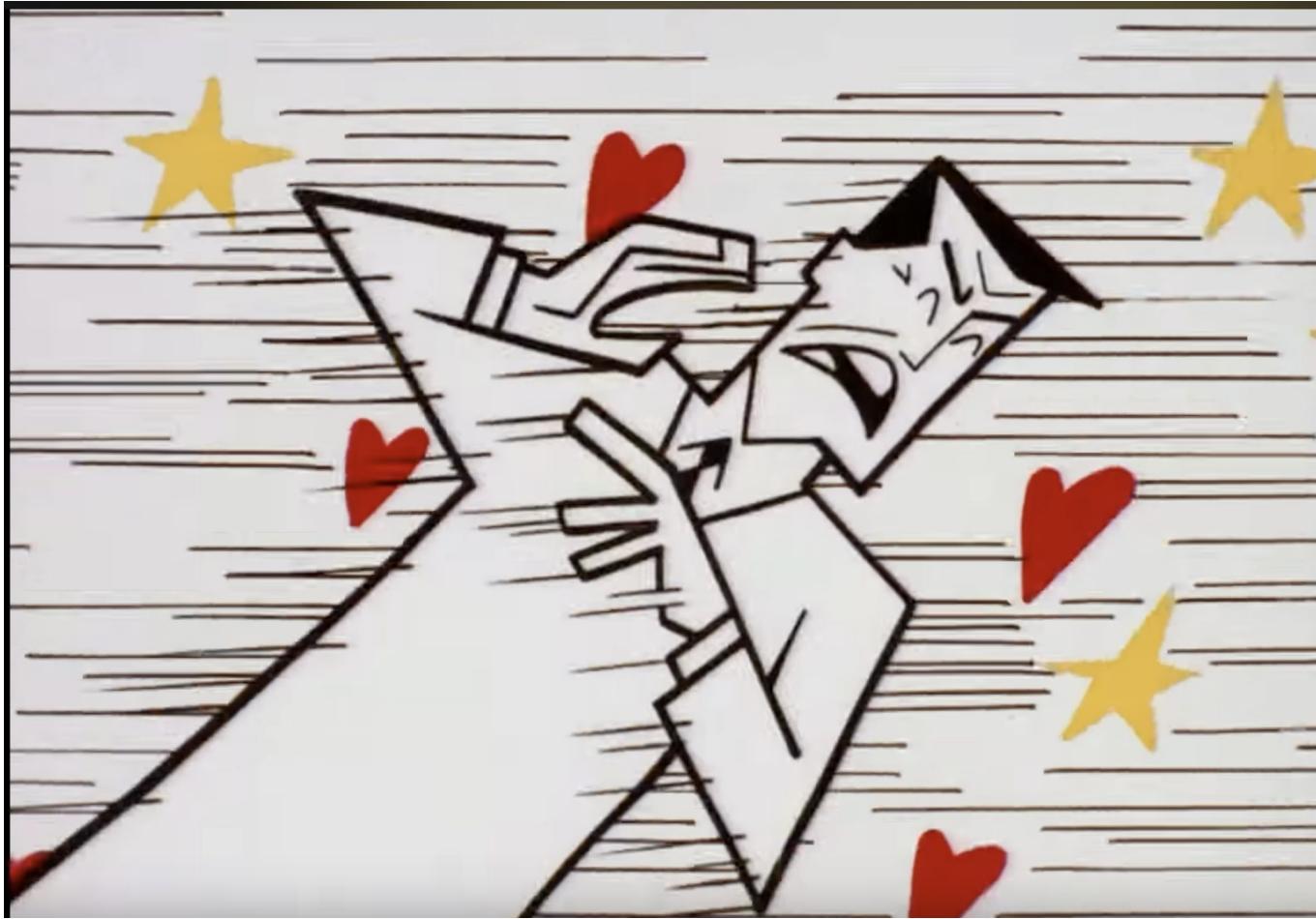
CONTRIBUTORS & MAINTAINERS



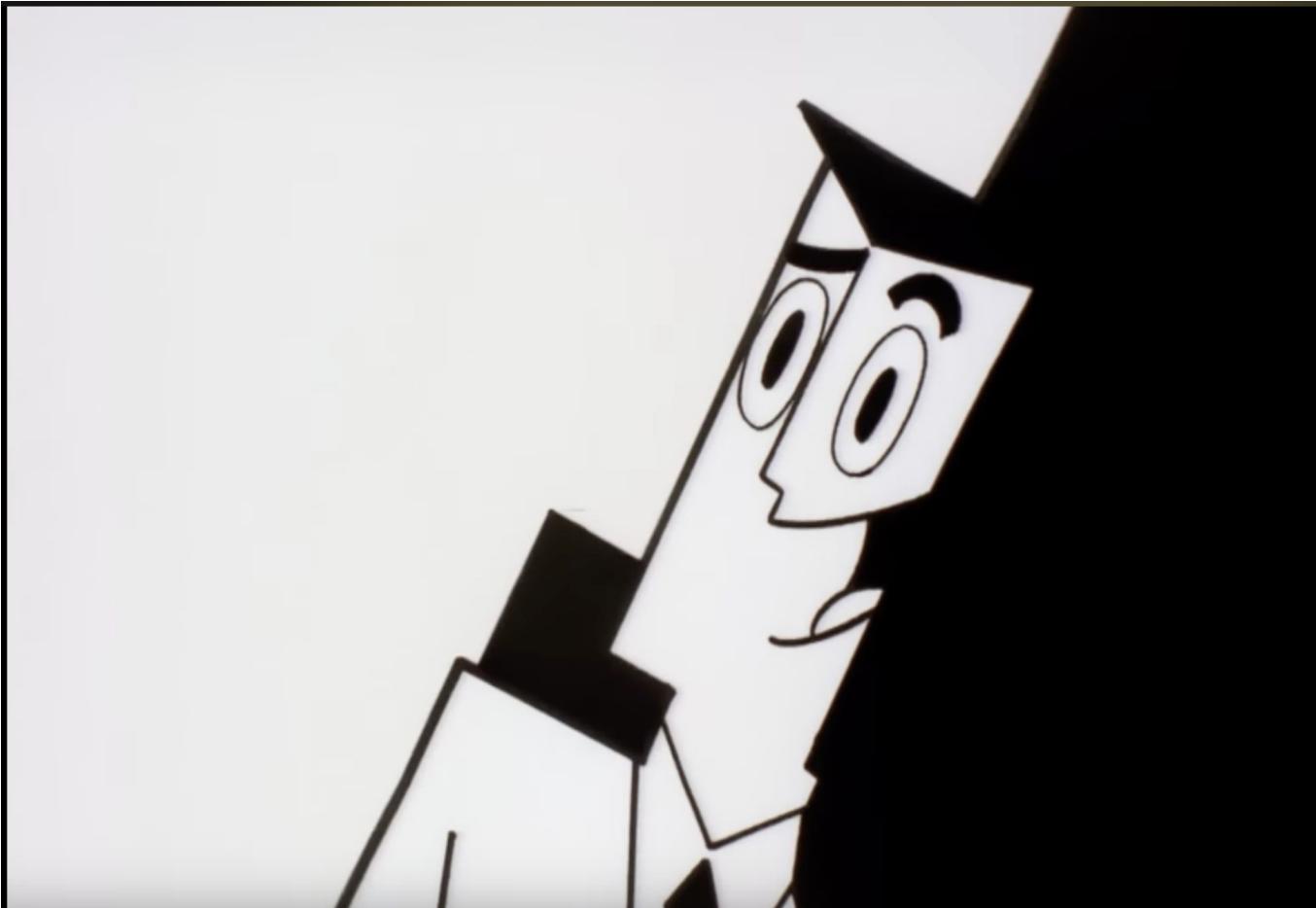


EASY TO USE

HACKABLE AND AGNOSTIC



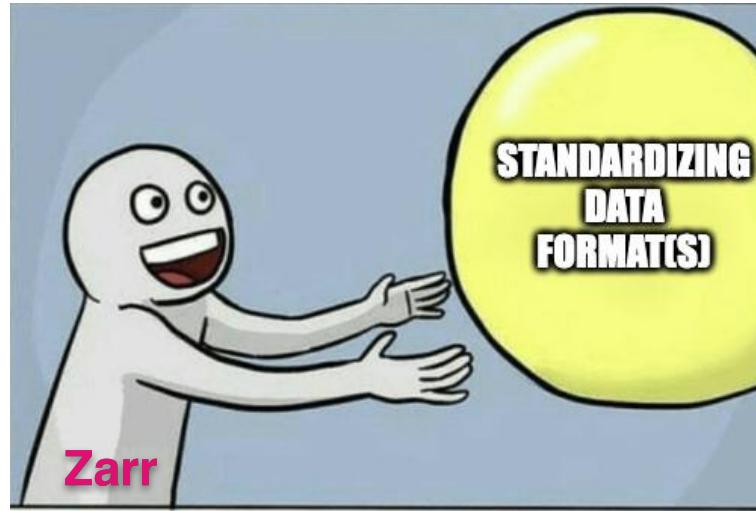






We have a large and diverse active community!

But... 



We needed a
structured way to
solicit and ***process***
the feedback!

zarr.dev/zeps



Zarr

Search ZEP

Zarr Homepage

active ZEPs / ZEP0000

home

active ZEPs

ZEP0000

draft ZEPs

template

implementations council

ZEP meetings

join the community

ZEP 0 — Purpose and process

Author: Sanket Verma (@MSanKeys963), Zarr

Email address: svsanketverma5@gmail.com

Status: Active

Type: Process

Created: 2022-14-03

Discussion: <https://github.com/zarr-developers/governance/pull/16>

What is ZEP?

ZEP stands for Zarr Enhancement Proposal. A ZEP is a design document providing information to the Zarr community, describing a modification or enhancement of the Zarr specification, a new feature for its processes or environment. The ZEP should provide specific proposed changes to the Zarr specification and a narrative rationale for the specification changes.

We intend ZEPs to be the primary mechanism for evolving the spec, collecting community input on major issues and documenting the design decision that has gone into Zarr. In addition, the ZEP author is responsible for building consensus within the community and documenting dissenting opinions.

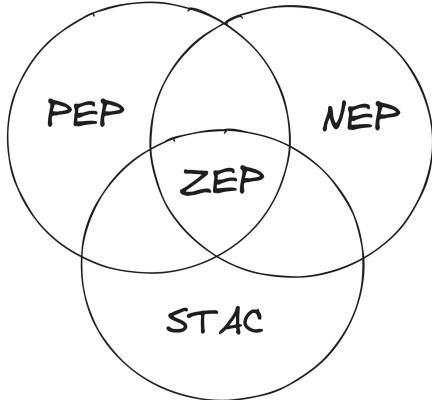
Because the ZEPs are maintained as text files in a versioned repository, their revision history is the historical record of the feature proposal.

WHERE:

- Developers refer to contributors and maintainers of the project
- User(s) refers to an individual or group of individuals or the broader community using the project in any way.

How we did it?





Lots of Reading



Previous Experience



Understanding the needs of the community

Braindump



ZEP Inception Blog post



How do we adopt a ZEP?





Demo

zarr.dev/implementations

 **Zarr**
chunked, compressed, N-dimensional arrays

Documentation Contribute Python Tutorial 

Content

ABOUT

Description
Applications
Features
Sponsorship
Videos

SUBPAGES

Adopters
Blog
Community
Conventions
Datasets
Implementations
Office Hours
Slides
Specification
ZEPs

Zarr Implementations

Zarr is a data storage format based on an open-source [specification](#), making implementations across several languages possible. It is used in various domains, including geospatial, bio-imaging, genomics, data science, and HPC.



Implementations are listed (in alphabetical order) as follows:

C	C++	Java	Javascript	Julia	Python	R	Rust
NetCDF-C	GDAL	JZarr	Zarr.js	Zarr.jl	Zarr-Python	Rarr	Rust-N5
		Tensorstore	N5-Zarr	Zarr-js	Zarrita	Pizzarr	Zarr
	Xtensor-Zarr	NetCDF-Java	Zarrita.js				Zarrs
		Z5					

52

zarr.dev/adopters

arr Adopters

If you're using Zarr in any way and would like to be added on this page, please drop your logo and blurb [here](#).

anks to the amazing community, Zarr is widely adopted and used by these groups. Here are the logos (in alphabetical order):

carbon)plan

Zarr is used by [CarbonPlan](#) as a storage format for analysis and visualization of climate data.

 COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

 COLUMBIA CLIMATE SCHOOL
LAMONT-DOHERTY EARTH OBSERVATORY

How does V3 data looks like?



WEBKNOSSOS

Citation:

Motta et al., Science 2019

Maintained by:



/normanrz

(Norman Rzepka)



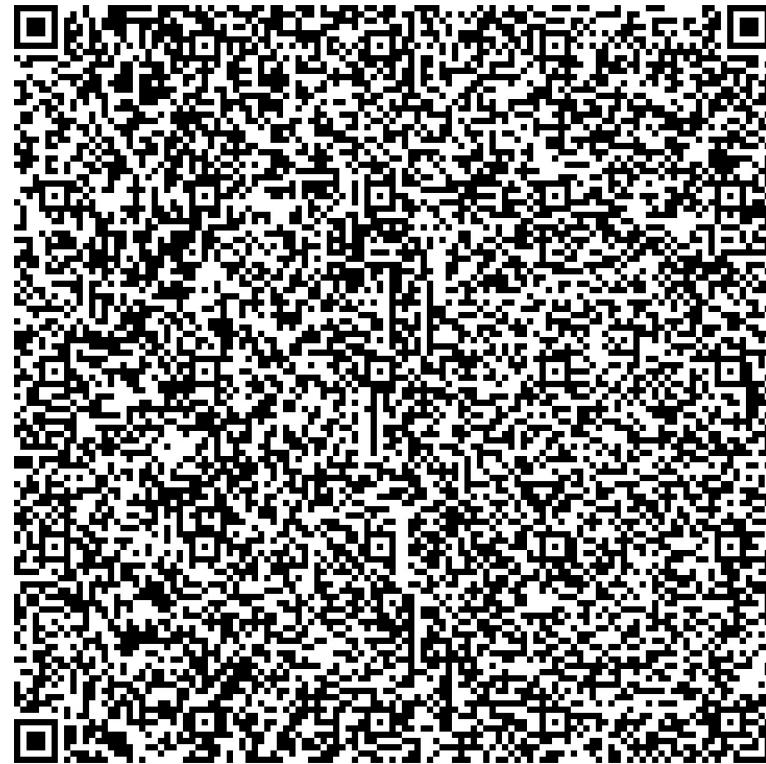
<https://webknossos.org/links/NxMfc4B65CmTVGhr>



<https://github.com/google/neuroglancer>

Maintained by:  /jbms

(Jeremy Maitin-Shephard)



[Neuroglancer visualiser](#)

Join us at community, ZEP, core-dev meetings or office hours!

Zarr Public Calendar

Today   **May 2024** 

 [Print](#)  [Week](#)  [Month](#)  [Agenda](#) 

Mon	Tue	Wed	Thu	Fri	Sat	Sun
29	30	1 May	2	3	4	5
		23:30 Zarr Communi	21:30 ZEP Meeting			
6	7	8	9	10	11	12
		20:30 GeoZarr Bi-we	21:30 Zarr Steering C			
		21:30 Zarr-Python R				
		23:30 Zarr Office Ho				
13	14	15	16	17	18	19
		20:30 GeoZarr Bi-we	21:30 ZEP Meeting	19:30 Zarr-Python D		
		23:30 Zarr Communi				
20	21	22	23	24	25	26
		23:30 Zarr Office Ho		03:30 Zarr-Python D		
27	28	29	30	31	1 Jun	2
		20:30 GeoZarr Bi-we	21:30 ZEP Meeting	19:30 Zarr-Python D		
		23:30 Zarr Communi				

Events shown in time zone: India Standard Time - Kolkata

 [Google Calendar](#)

<https://zarr.dev/community-calls/>

Thank you!

