MINISTÈRE
DE L'ENSEIGNEMENT
SUPÉRIEUR
ET DE LA RECHERCHE
Liberté
Égalité
Fraternité

Ouvrir la science !

Software Heritage
THE GREAT LIBRARY OF SOURCE CODE

Software as Infrastructure and Infrastructures for Software

Roberto Di Cosmo, co-chair Software and Source Code College
Director, Software Heritage, Inria and Université Paris Cité

@rdicosmo
roberto@dicosmo.org

MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE
Liberté Égalité Fraternité

Ouvrir la science !

Software Heritage
THE GREAT LIBRARY OF SOURCE CODE

Software is not data, it is **executable knowledge**

Apollo 11 source code (excerpt)

```
P63SPOT3        CA      BIT6            # IS THE LR ANTENNA IN POSITION 1 YET
                EXTEND
                RAND    CHAN33
                EXTEND
                BZF     P63SPOT4        # BRANCH IF ANTENNA ALREADY IN POSITION 1

                CAF     CODE500         # ASTRONAUT:    PLEASE CRANK THE
                TC      BANKCALL        #               SILLY THING AROUND
                CADR    GOPERF1
                TCF     GOTOPOOH        # TERMINATE
                TCF     P63SPOT3        # PROCEED       SEE IF HE'S LYING

P63SPOT4        TC      BANKCALL        # ENTER         INITIALIZE LANDING RADAR
                CADR    SETPOS1

                TC      POSTJUMP        # OFF TO SEE THE WIZARD ...
                CADR    BURNBABY
```

Quake III source code (excerpt)

```
float Q_rsqrt( float number )
{
    long i;
    float x2, y;
    const float threehalfs = 1.5F;

    x2 = number * 0.5F;
    y = number;
    i = * ( long * ) &y; // evil floating point bit level hacking
    i = 0x5f3759df - ( i >> 1 ); // what the fuck?
    y = * ( float * ) &i;
    y = y * ( threehalfs - ( x2 * y * y ) ); // 1st iteration
// y = y * ( threehalfs - ( x2 * y * y ) ); // 2nd iteration, this
can be removed

    return y;
}
```

Len Shustek, Computer History Museum 2006

"Source code provides a view into the mind of the designer."

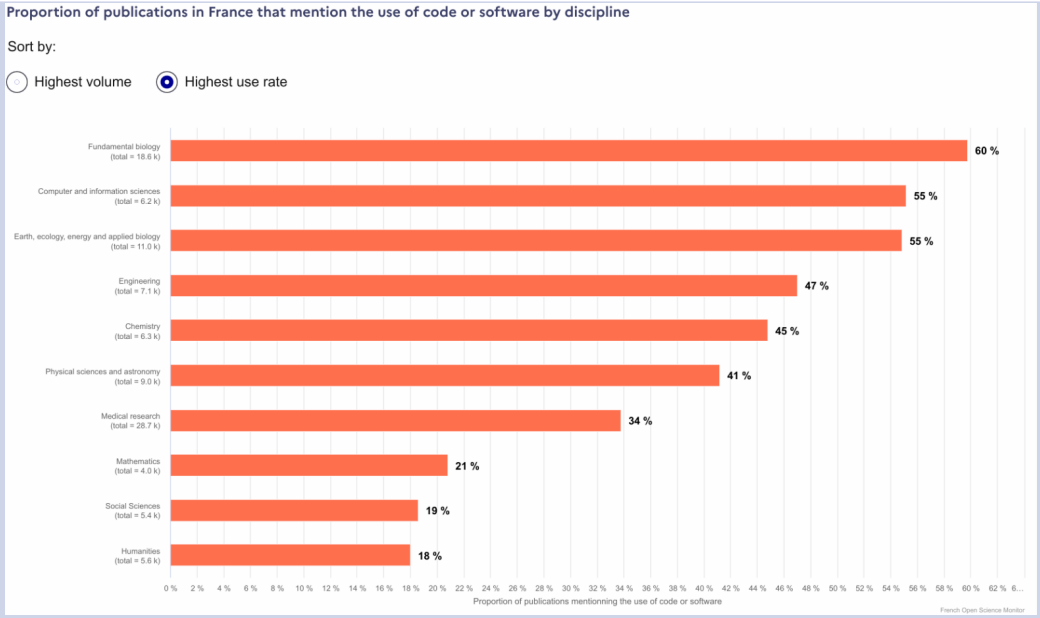
MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE
Liberté Égalité Fraternité

Ouvrir la science !

Software Heritage
THE GREAT LIBRARY OF SOURCE CODE

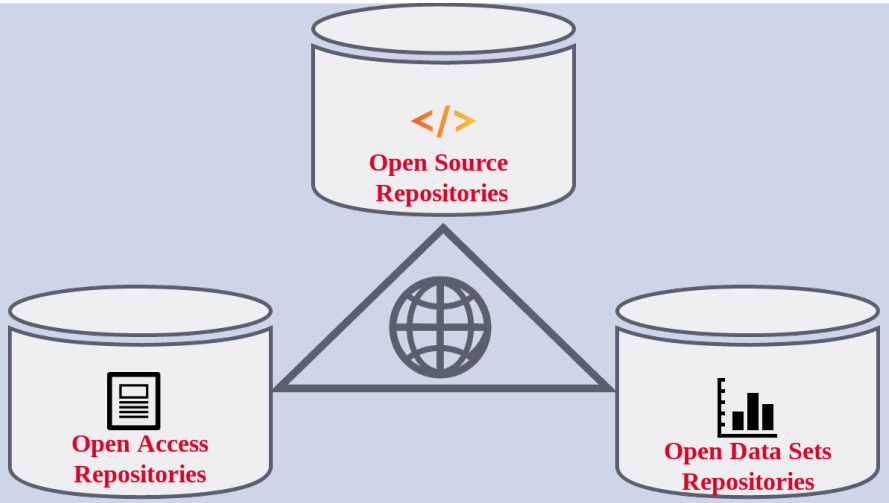
Software is a key Infrastructure for Science

Software powers modern research



Across ***all*** disciplines

Software : a pillar of Open Science



The links in the picture are *essential*

Software may be a **tool**, an **outcome** and a **research object**

Open Source (open access **to the source code**) is **necessary**

MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE
Liberté Égalité Fraternité

Ouvrir la science !

Software Heritage
THE GREAT LIBRARY OF SOURCE CODE

We need Infrastructures for Software !

Technical Infrastructures	Organizational Infrastructures

• **Monitor software development and use**
• **Support software development**
• **Guarantee long term availability**
• **Ensure integrity and persistent identification**
• **Enable curation, reuse and citation**
• Support compliance with best practices and forthcoming legal obligations (CRA)
• Ease development and reuse (e.g. LLMs)
	• *Recognize software development on par with publications*
	• *Fund software maintenance*
	• *Support researchers with*
	  • *Licensing and tech transfer*
	  • *Curation and archival*
	  • *Reproducibility*
	• *and much more*

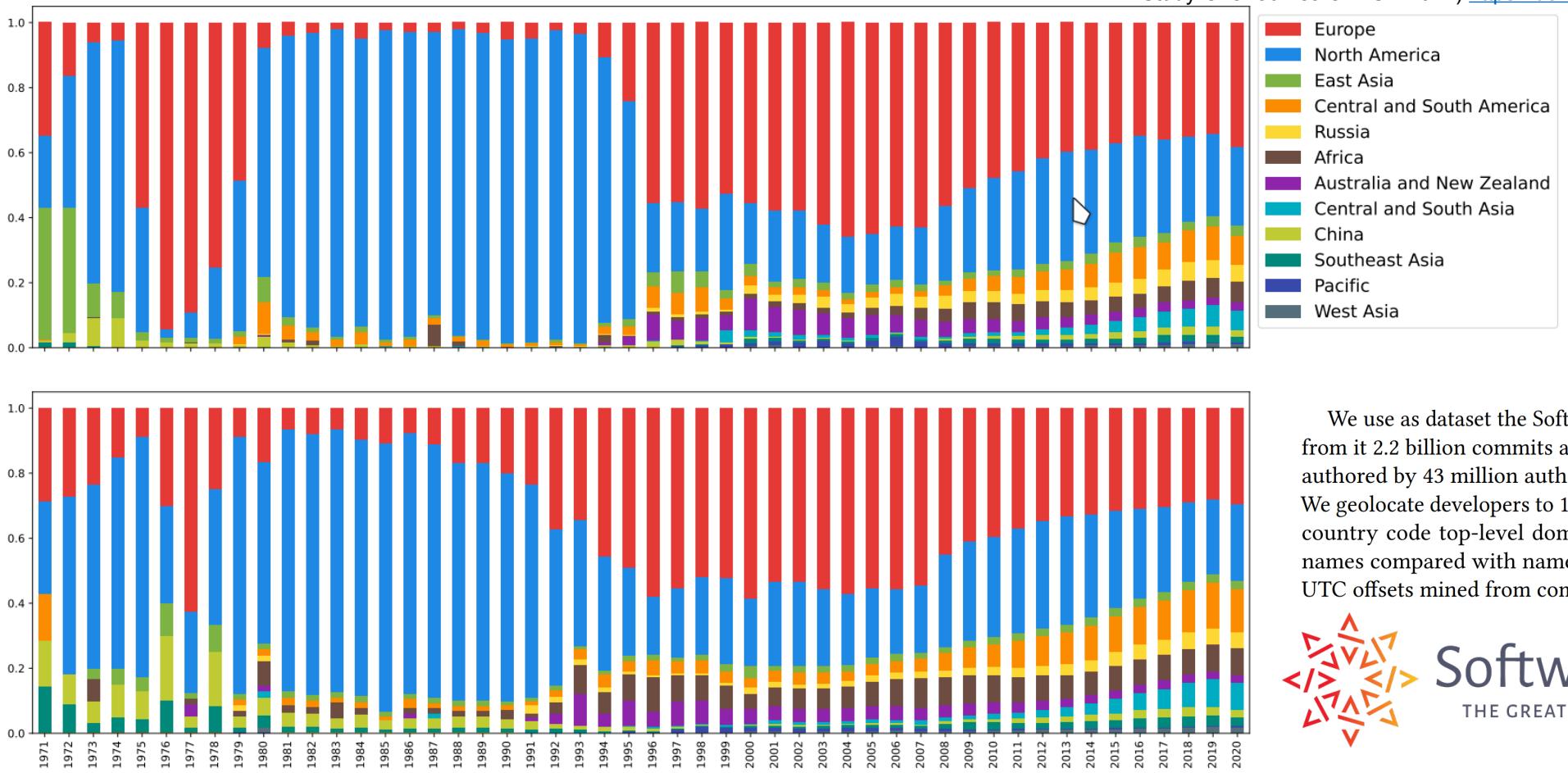
This session focuses on **technical infrastructures for software that are :**
Created or started in France, state of the art, and scalable european- and world-wide

Because software is naturally international !

MSR '22, May 23–24, 2022, Pittsburgh, PA, USA

Davide Rossi and Stefano Zacchiroli

Geographic Diversity in Public Code Contributions: An Exploratory Large-Scale Study Over 50 Years. MSR 2022) https://doi.org/10.1145/3524842.3528471



We use as dataset the Software Heritage archive [3] and analyze from it 2.2 billion commits archived from 160 million projects and authored by 43 million authors during the 1971–2021 time period. We geolocate developers to 12 world regions, using as signals email country code top-level domains (ccTLDs) and author (first/last) names compared with name distributions around the world, and UTC offsets mined from commit metadata.

Software Heritage
THE GREAT LIBRARY OF SOURCE CODE

Figure 3: Ratio of commits (above) and active authors (below) by world zone over the 1971–2020 period.

Monitoring

The French Open Science Monitor: an open tool for monitoring public policy
- Eric Jeangirard (MESR)

Development

- Daniel Le Berre, Collège codes sources et logiciels du Comité pour la Science Ouverte, Centre de Recherche en Informatique de LENS

Archive, Reference, Describe, Cite

Software Heritage: A universal source code archive serving the scholarly ecosystem
- Morane Gruenpeter, Software Heritage

Archiving, referencing and describing software with HAL and Software Heritage
- Yannick Barborini, Centre pour la Communication Scientifique Directe