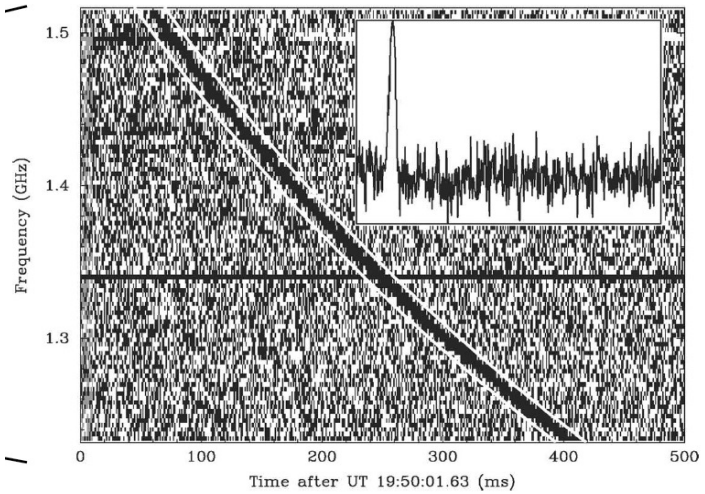
Motivation: In the past, we could make discoveries with archives...

"Fast radio bursts"

Lorimer et al. (2007)



REPORTS

A Bright Millisecond Radio Burst of Extragalactic Origin

D. R. Lorimer,[1,2*] M. Bailes,[3] M. A. McLaughlin,[1,2] D. J. Narkevic,[1] F. Crawford[4]

Pulsar surveys offer a rare opportunity to monitor the radio sky for impulsive burst-like events with millisecond durations. We analyzed archival survey data and found a 30-jansky dispersed burst, less than 5 milliseconds in duration, located 3° from the Small Magellanic Cloud. The burst properties argue against a physical association with our Galaxy or the Small Magellanic Cloud. Current models for the free electron content in the universe imply that the burst is less than 1 gigaparsec distant. No further bursts were seen in 90 hours of additional observations, which implies that it was a singular

ratios greater than 4 with the use of a matched filtering technique (7) optimized for pulse widths in the range 1 to 1000 ms. The burst was detected in data taken on 24 August 2001 with DM = 375 cm^{-3} pc contemporaneously in three neighboring beams (Fig. 1) and was located ~3° south of the center of the Small Magellanic Cloud (SMC).

The pulse exhibited the characteristic quadratic delay as a function of radio frequency (Fig. 2) expected from dispersion by a cold ionized plasma along the line of sight (δ). Also evident was a significant evolution of pulse width across the observing frequency band. The behavior we observed, where the pulse width W

Astron. Soc. **417**, 2982 (2011). **398**, 1858 (2009). 10.1126/science.1234209

A Population of Fast Radio Bursts at Cosmological Distances

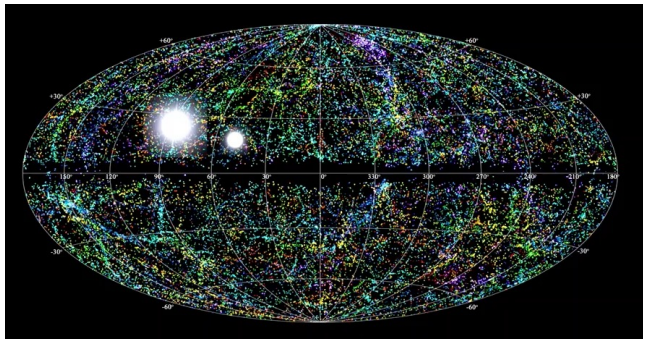
D. Thornton,[1,2*] B. Stappers,[1] M. Bailes,[3,4] B. Barsdell,[3,4] S. Bates,[5] N. D. R. Bhat,[3,4,6] M. Burgay,[7] S. Burke-Spolaor,[8] D. J. Champion,[9] P. Coster,[2,3] N. D'Amico,[10,7] A. Jameson,[3,4] S. Johnston,[2] M. Keith,[2] M. Kramer,[9,1] L. Levin,[5] S. Milia,[7] C. Ng,[9] A. Possenti,[7] W. van Straten[3,4]

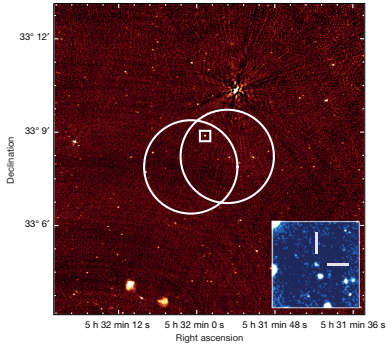
Searches for transient astrophysical sources often reveal unexpected classes of objects that are useful physical laboratories. In a recent survey for pulsars and fast transients, we have uncovered four millisecond-duration radio transients all more than 40° from the Galactic plane. The bursts' properties indicate that they are of celestial rather than terrestrial origin. Host galaxy and intergalactic medium models suggest that they have cosmological redshifts of 0.5 to 1 and distances of up to 3 gigaparsecs. No temporally coincident x- or gamma-ray signature was identified in association with the bursts. Characterization of the source population and identification of host galaxies offers an opportunity to determine the baryonic content of the universe.

measure (DM) is related to the integrated colu
density of free electrons along the line of si
to the source and is a proxy for distance. T
frequency-dependent pulse broadening occ
as an astrophysical pulse is scattered by an
homogeneous turbulent medium, causing a ch
acteristic exponential tail. Parameterizing
frequency dependence of δt and W as α and
respectively, we measured $\alpha = -2.003 \pm 0.0$
and $\beta = -4.0 \pm 0.4$ for FRB 110220 (Table 1
Fig. 2), as expected for propagation throug
cold plasma. Although FRB 110703 shows no
dence of scattering, we determined $\alpha = -2.00$
0.006. The other FRBs do not have suffic

[1]Jodrell Bank Centre for Astrophysics, School of Physics
Astronomy, University of Manchester, Manchester M13
UK. [2]Commonwealth Science and Industrial Research Org

- One burst every 10 seconds!
- Found in archival data in 2006 in 5-year old data!
- Now: new field – all radio telescopes now going after them: cosmological importance
- Caused our on-line processing to be changed: online-buffer, polarization capture etc.
- Aim: online recognition and trigger
- False-alarm threshold important





LETTER

doi:10.1038/nature20797

A direct localization of a fast radio burst and its host

S. Chatterjee[1], C. J. Law[2], R. S. Wharton[1], S. Burke-Spolaor[3,4,5], J. W. T. Hessels[6,7], G. C. Bower[8], J. M. Cordes[1], S. P. Tendulkar[9], C. G. Bassa[6], P. Demorest[3], B. J. Butler[3], A. Seymour[10], P. Scholz[11], M. W. Abruzzo[12], S. Bogdanov[13], V. M. Kaspi[9], A. Keimpema[14], T. J. W. Lazio[15], B. Marcote[14], M. A. McLaughlin[4,5], Z. Paragi[14], S. M. Ransom[16], M. Rupen[11], L. G. Spitler[17] & H. J. van Langevelde[14,18]

Motivation: ...but this may not be longer possible

- Previously: (most) data could be stored and reanalyzed/reused in data life cycle
- Now: data rates increasingly large – even after on-line analysis (FPGA,GPU,HPC,AI,...) – only parts of data can be stored!
- Soon: only tiny fraction of data can be stored: **dramatic loss of information!**

20 TB/night
1.5 PB/yr

40 TB/night
4 PB/yr

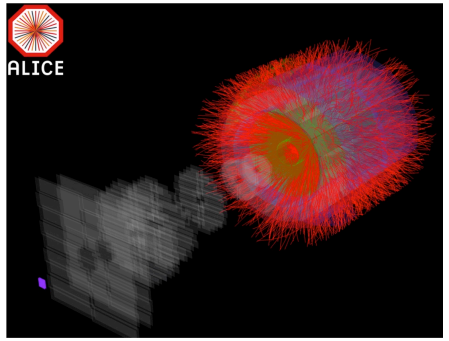
13 EB/day
300 PB/yr

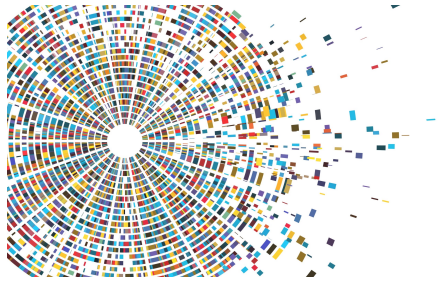
2 PB/day
4 PB/yr

Making decisions on the fly, what is archived or not – archives need to capture decisions!

Decisions also important when triggering other telescopes ("multi-messenger").

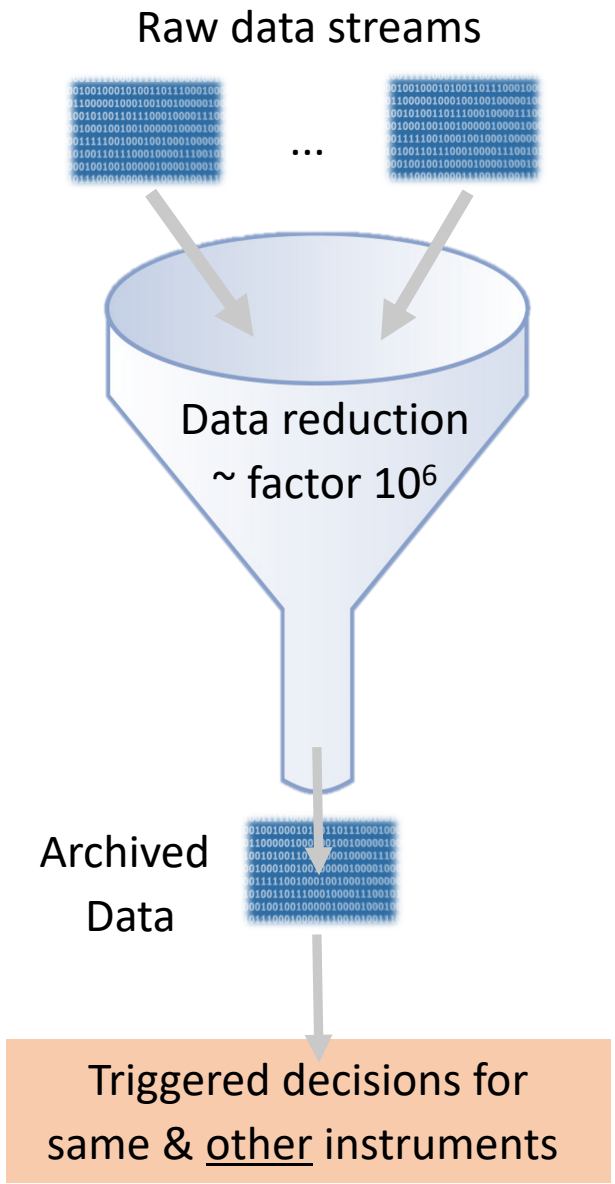
A general challenge with modern experiments & beyond



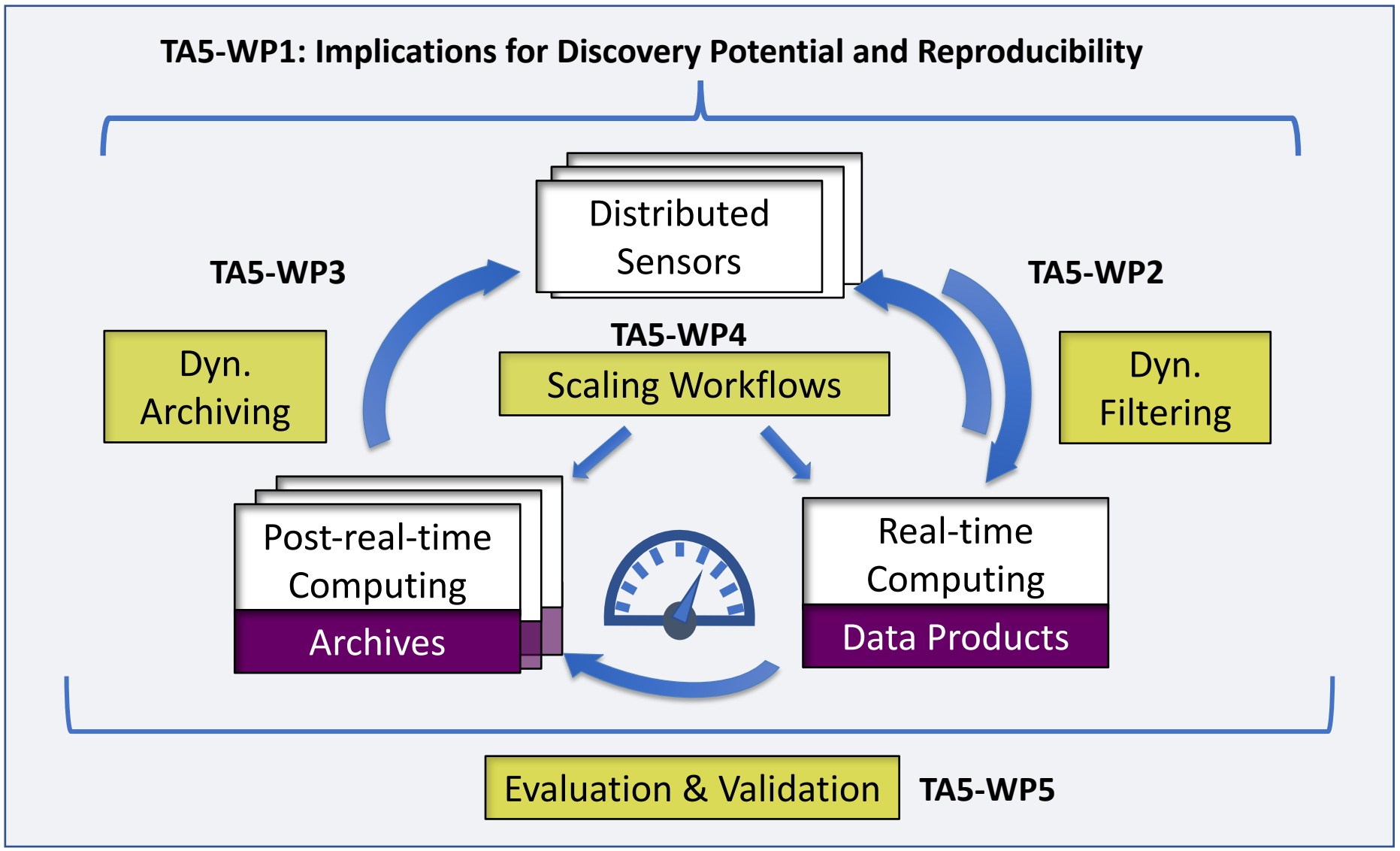


- Data rates are increasingly large – even after on-line analysis (FPGA, GPU, HPC, AI, ML etc.) only small parts of data can be stored
- Result: dramatic loss of information by irreversible loss of data
- Already reality in e.g. parts of astronomy & HEP!
- Strategy: real-time dynamical filtering, dynamical archives & scalability – this is already a challenge on its own.
- But, even more, what are the implications for reproducibility, discovery potential & interpretation?
- What to store? How to select? Minimizing impact of climate?

Many far-reaching technical questions:
- Metadata in the context of real-time data analysis
- Effective data reductions and GreenIT



Adressing these problems in TA 5: „Data Irreversibility“



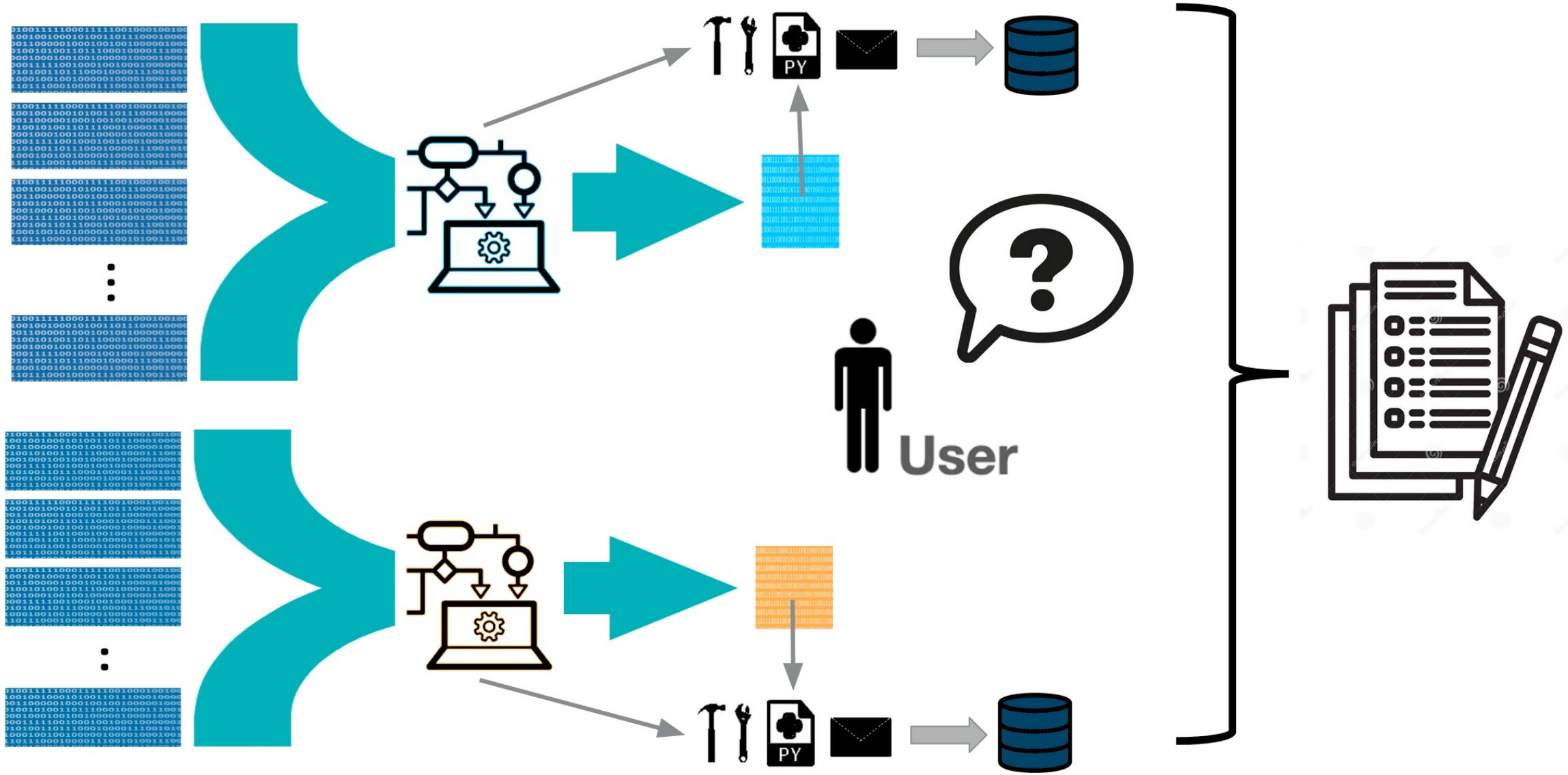
Blueprint for other fields of data-intensive science

Dynamic Life Cycle Model

Challenge: verification in light of (almost) no raw data in archives

WP 1 Implications for discovery potential and reproducibility

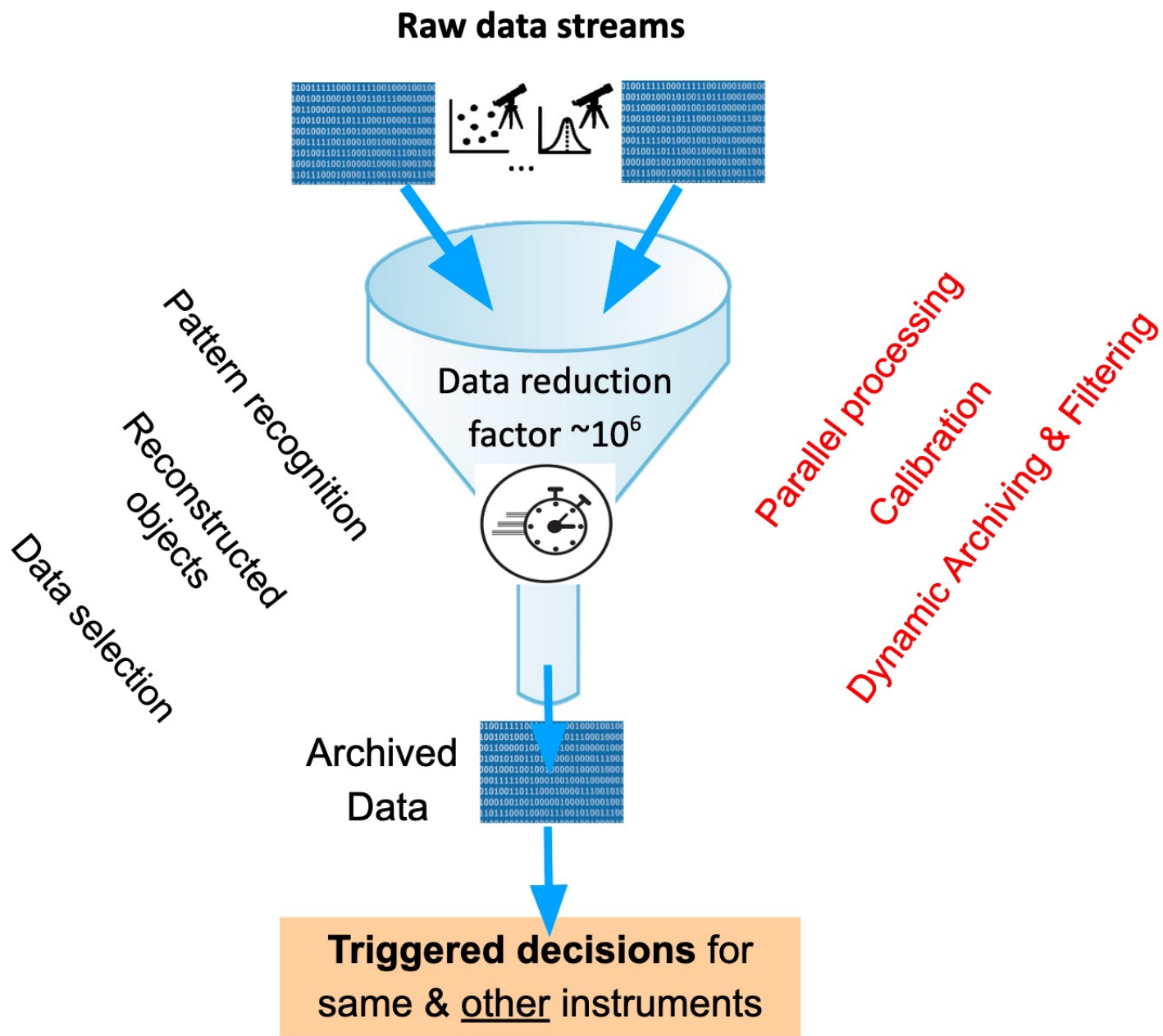
Addressing: tension and interplay between reproducibility of filtered data and implications for discovery potential.
Delivering: possible solutions, curation criteria and structural requirements on metadata in light of FAIR principles.



WP2 Dynamic filtering

Addressing: Discarding irrelevant information with minimal time budget using filtering of various forms of noise and irrelevant backgrounds

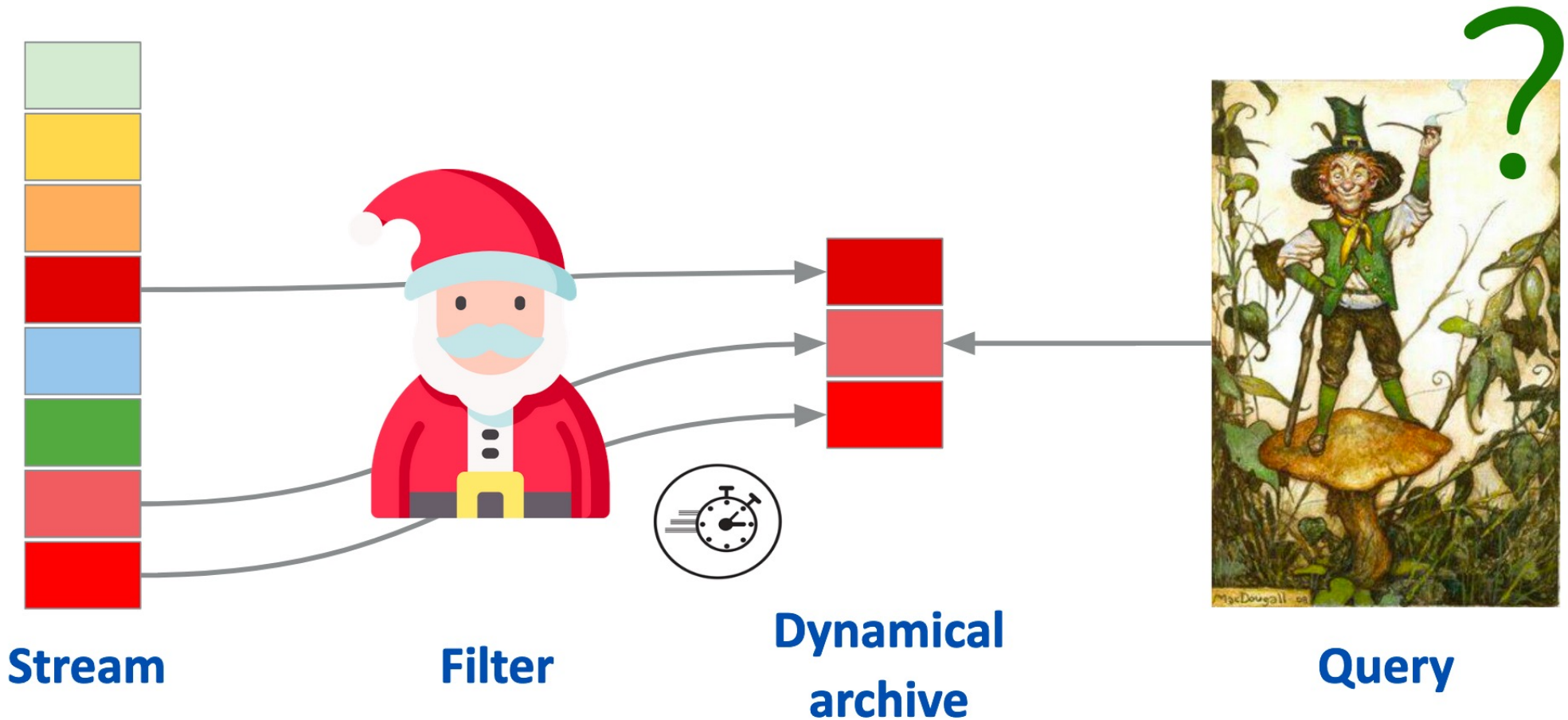
Delivering: Solutions for real-time selection of (anomalous) signals, including a description of decisions in metadata



WP3 Dynamic archiving

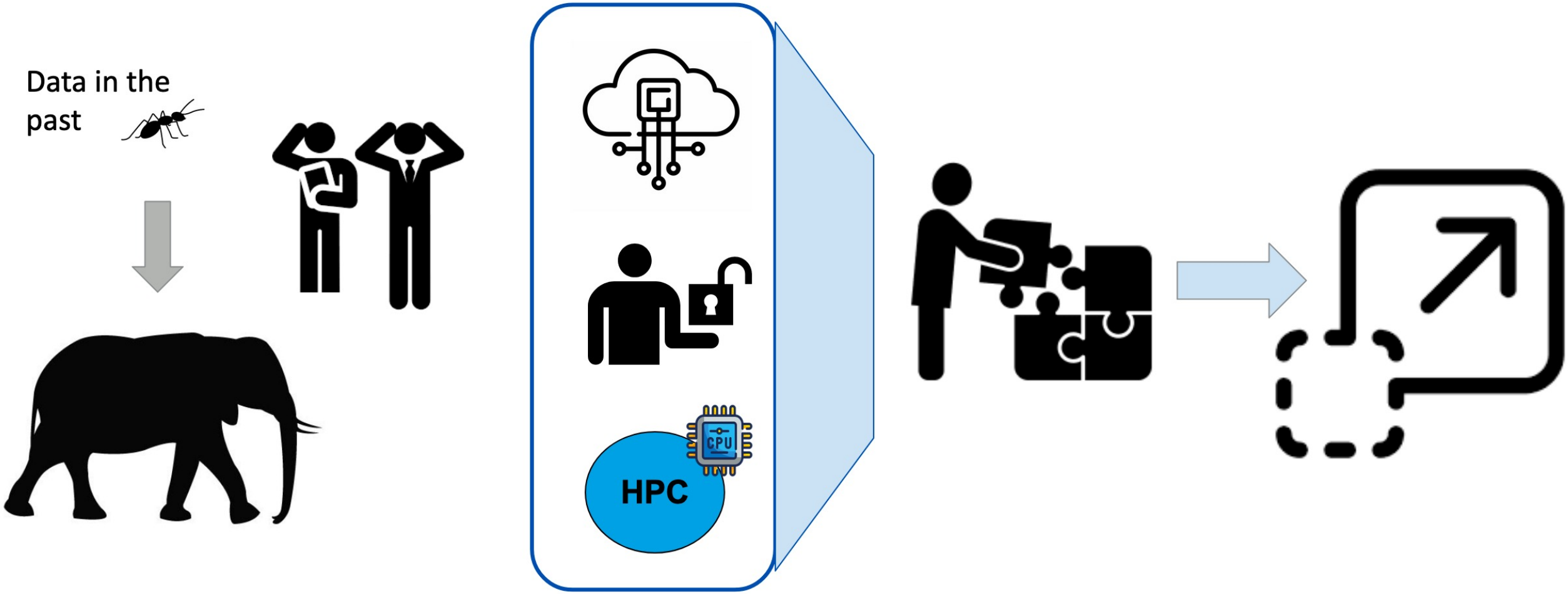
Addressing: Archives will be systematically incomplete: selection depended on (some past) real-time selection that needs to be tuned and updated to optimize future real-time processes (WP2).

Delivering: Methods for queries of (multiple) dynamic archives which returns: 1) collection of real-time selections that would have been made, 2) estimate of how well archived datasets represent real-time data stream for specific query (WP1+5).



WP4 Scaling workflows

Addressing: Optimal use of hardware and software resources when analysing single huge data sets
Delivering: Report on technology solutions for scaling the ”online” and ”offline” workflows



Data today/soon

Small scale solutions vs proven and new technology

Combine & optimise

Scale

WP5 Evaluation and validation of instrument response & characteristics

Addressing: data quality assurance for systems with irreversible data processing in presence of variable, unpredictable background

Delivering: anomaly detection and predictive maintenance for real-time systems exploiting machine learning solutions with on-line feedback to archive metadata and studyimpact on noise