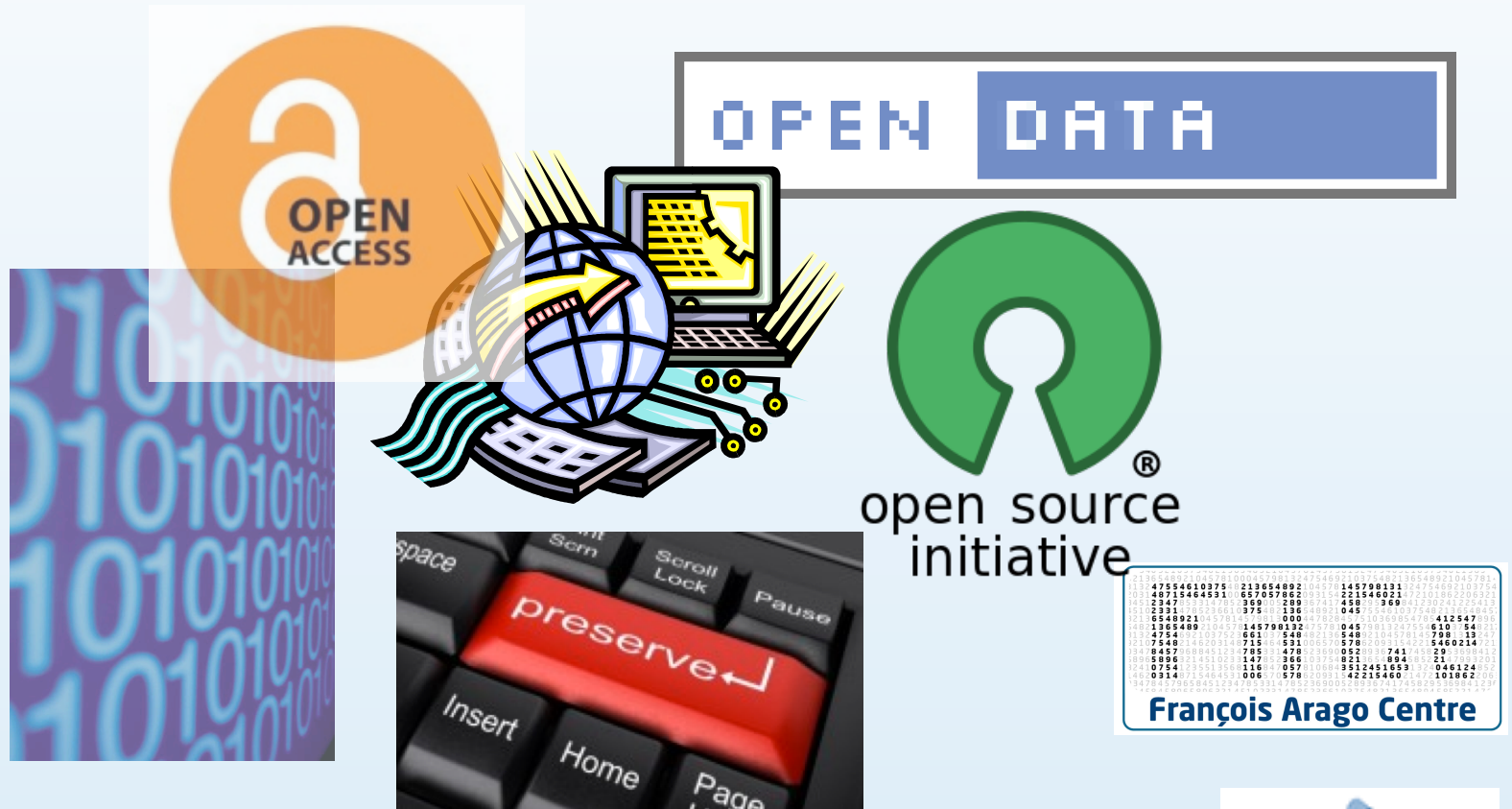


Access et préservation des données, des solutions pour tous ?



Volker Beckmann
François Arago Centre / APC Paris





Outline



- Open data / open source / open access motivation
(see also Research Data Alliance presentation by Vincent at noon)
- Concerns
- Data preservation: data + s/w + knowledge
- Challenges
- Actions (?)





Motivation open data



- Get most science / return out of projects
 - Unforeseen use of data
 - Stimulate science through competition
 - Improvement of analysis techniques
 - Credibility of scientific results
 - Support scientists / institutes in less fortunate setting
-
- Funding agencies request open access/data/...
 - ANR, EC (H2020), ESA, ...

Open Access et gestion des données dans Horizon 2020

*Intégrer les
nouveaux attendus
dans vos projets de
recherche*



**29 et 30 juin
2015**

Université Paris Descartes
12 rue de l'École de Médecine
75005 Paris

Métro lignes 4 et 10



UNIVERSITÉ
PARIS DESCARTES

PARIS
DESCARTES



Documentaire financé par le Fonds de la Recherche en Santé
Horizon 2020 - Programme de recherche et d'innovation
Horizon 2020 - Programme de recherche et d'innovation

**Avec le soutien de la
Commission
Européenne**

This project has
received funding
from the European
Union's Seventh
Framework Programme for
research, technological
development and
demonstration under grant
agreement no 612425.





Concerns



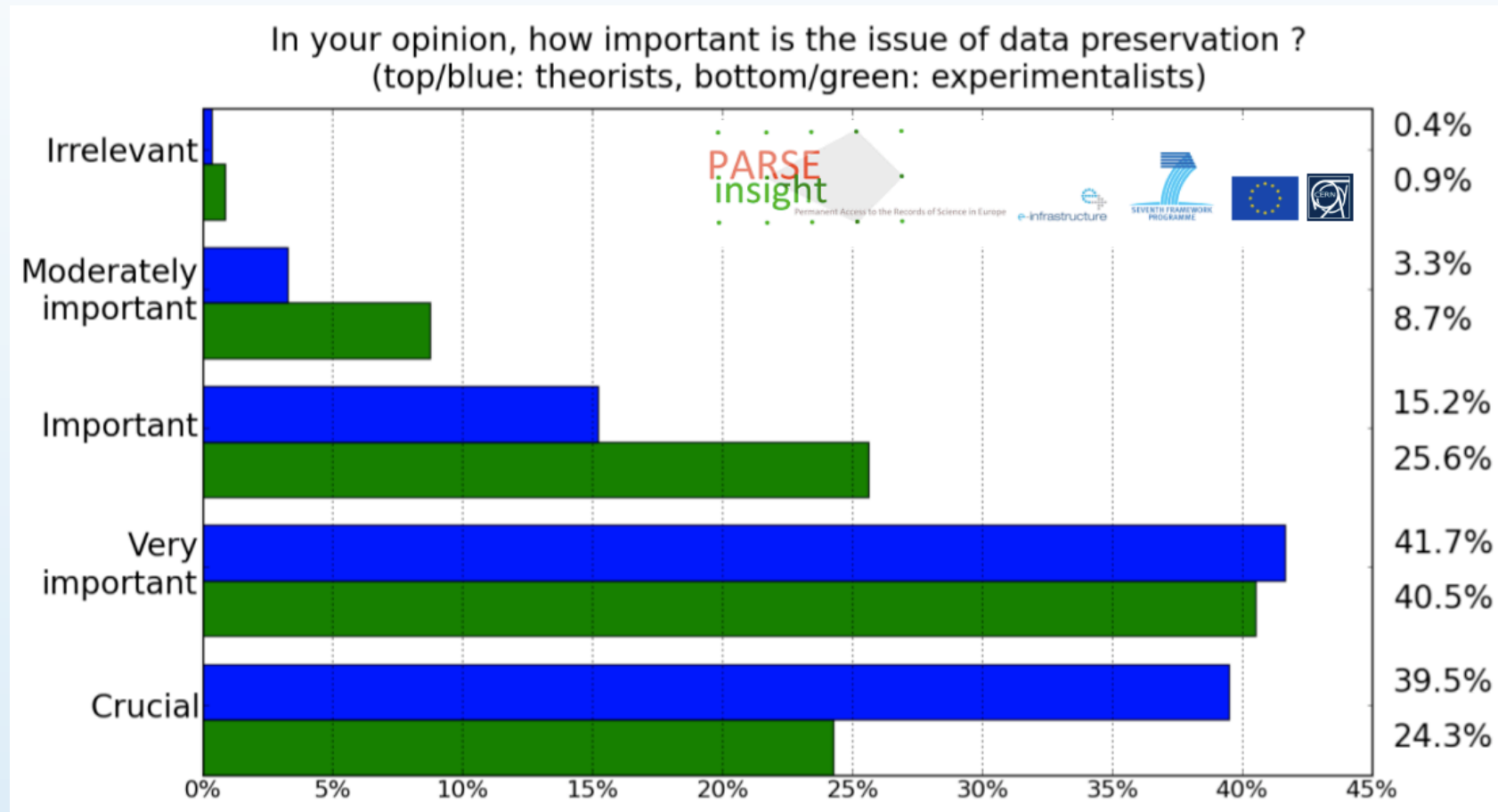
- “We do the work, somebody else does the science!”
→ Main expertise remains with consortium who built experiment
- “Open data leads to false results”
→ Also true for teams sometimes. Peer review.
- “Funding depends on exclusive data rights”
→ Changing landscape!
- “Open data/source requires user support”
→ true, but leads to visibility, feedback, improvement



Data preservation



Data are valuable beyond the end of a projects' life time



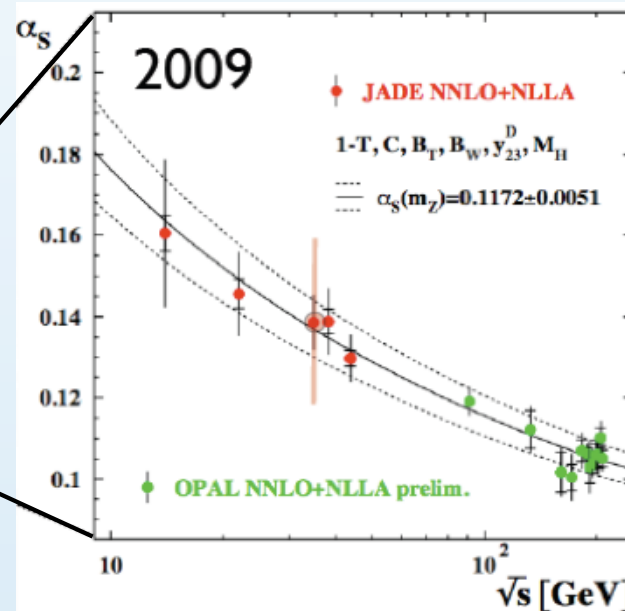
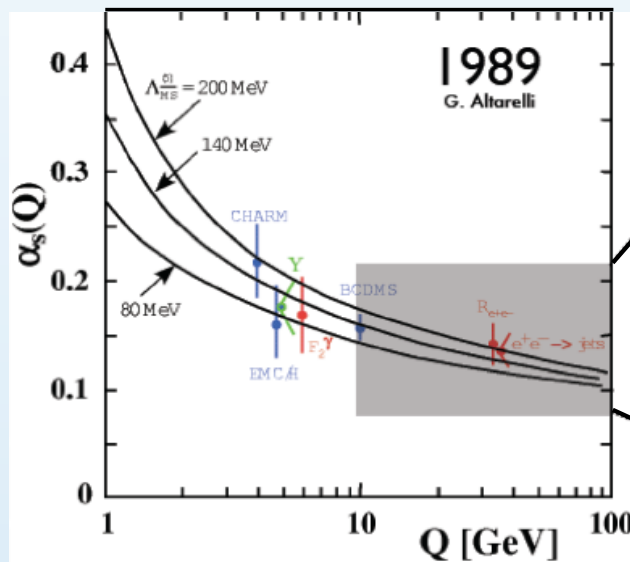
Holzner, Igo-Kemenes & Mele (2009):

« First results from the PARSE.Insight project: HEP survey on data preservation, re-use and (open) access », arXiv:0906.0485

Rescued data used for fundamental results

- JADE experiment
 - Data rescued “in extremis”
 - Stored by a scientist in a suitcase
 - Re-analysed 20 years later

10 publications



How many of these stories did **not** have a happy end?



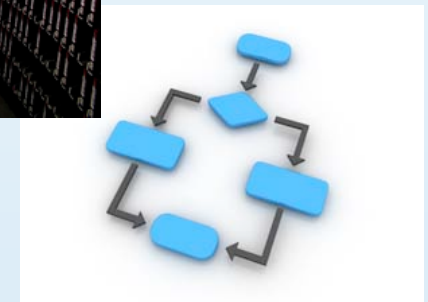
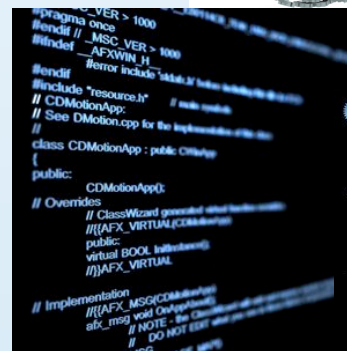
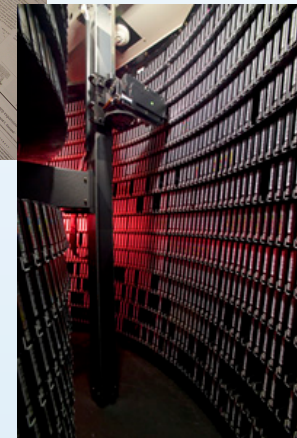
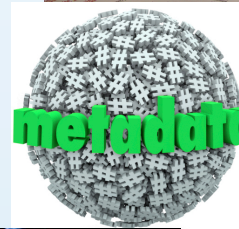
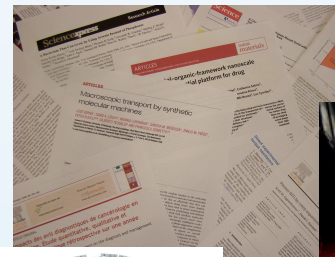
Data preservation



- Data are valuable beyond the end of a projects' life time

What needs to be preserved?

- Publications
- Documentation
- Raw and processed data
- Meta data
- Workflows
- Software
- Knowledge





Challenges



What needs to be preserved?

- Publications
 - Documentation
 - Raw and processed data
 - Meta data
 - Workflows
 - **Software**
- virtualisation
- **Knowledge**



Challenges



- Cost: preservation team not (necessarily) benefitting from preserved data
- Motivation
- Find common solutions / interfaces / standards
- Top-down approach?



Interfaces



Physics Data Repositories

- [HEP Data](#) – high-energy physics reaction database of Numerical H
- [National Nuclear Data Center](#) – includes nuclear structure, reaction
- [Neutron Monitor Database](#) – worldwide network of neutron monitor
- [GeV energy range.](#)
- [NIST Physical Standards Laboratory](#) – physical reference data and
- [SuperDARN \(Dual Auroral Radar Network\)](#)
- [SuperMAG \(ground based magnetometer data\)](#)
- [Virtual Modeling Repository \(simulations and simulated output\)](#)

Astronomy Data Repositories

- [HEASARC](#) – NASA's High Energy Astrophysics Science Archive Research Center
- [Heliophysics Integrated Observatory \(HELIO\)](#)
- [Infrared Science Archive](#) – NASA's science and data center for infrared astronomy
- [International Planetary Data Alliance](#) Solar system (planets, etc.)
- [NASA's Earth Observing System Data and Information System](#) – A distributed network of data systems responsible for data archival and distribution.
- [NASA Extragalactic Database](#) – NASA's archive of data for over 3 million extragalactic objects
- [National Space Science Data Center](#) – Archive for NASA space mission data
- [NOAA National Geophysics Data Center](#)
- [Planetary Data System](#) – organized into 'nodes' ...imaging, plasma, atmospheres, rings, small bodies, etc.
- [Sloan Digital Sky Survey](#) – Download optical images of the sky. See also, [SkyServer](#) for educational portal to the data.
- [Solar Data Analysis Center](#) – Based at Goddard Space Flight Center, NASA, they will archive data that's of interest to the community.
- [Southwest Data Display and Access System](#)
- [Space Physics Interactive Data Resource](#)
- [Space Physics Archive Search & Extract](#) – Metadata standards for describing space physics archives
- [Virtual Energetic Particle Observatory](#)
- [Virtual Heliophysics Observatory](#)
- [Virtual Ionosphere Thermosphere Mesosphere Observatory \(upper atmosphere\)](#)
- [Virtual Magnetospheric Observatory](#)
- [Virtual Radiation Belt Observatory](#)
- [Virtual Space Physics Observatory](#) – Catalog of space physics data , archives, instruments, etc.

Other efforts that are more IR-ish (just data from one institution)

- [Virtual Solar-Terrestrial Observatory](#)
- [Virtual Solar Observatory](#) – mostly remote-sensing images, some spectra
- [Virtual Wave Observatory](#) also has spectra

National Astronomy Repositories

- [International Virtual Observatory Alliance\(IVOA\)](#) – Defines metadata standards and XML schema for astronomical data documentation. View each member observatory by country at <http://www.ivoa.net/pub/members/>
- [US National Virtual Observatory](#) – Deposit/Access to astronomical data, including ground and space-based telescopes based in the US. Includes data analysis tools



Conclusions



- Scientific productivity benefits from open data / open source / scientific data preservation
- Cost & motivation
- Software preservation: virtualisation
- Knowledge / expertise preservation?





Conclusions + Actions?



- Scientific productivity benefits from open data / open source / scientific data preservation
 - Cost & motivation
 - Software preservation: virtualisation
 - Knowledge / expertise preservation?
-
- Should CNRS (IN2P3) push harder for projects to have open data/source policy and concepts for scientific data preservation?
 - Do we need a centralised s/w virtualisation/preservation system?

