



Results from the PARSE.Insight project

A survey on HEP data preservation, re-use and (open) access

<http://arxiv.org/abs/0906.0485>

- Background information
- The HEP Survey
- Conclusions

Andre Holzner (FP7/CERN), Peter Igo-Kemenes (Gjøvik/CERN), Salvatore Mele (CERN)

CERN Symposium on Data Preservation and Long Term Analysis in HEP
CERN - December 7th, 2009

Actors

← Alliance for Permanent Access →

Group of stakeholders in digital preservation of data and publications.

Purpose: best practices and dialogue with the EC

- Research (CERN, ESA, STFC, Helmholtz, MPG)
- National Libraries (Germany, Netherlands, UK)
- Publishers (STM association)

CERN involvement through
Scientific Information Service/Open Access



Salvatore Mele | CERN | Dec. 7, 2009



Actors



e-Infrastructure = connectivity (GEANT) + Grid (EGEE)

Other activities = repositories (i.e. DB for publications)

Possible future = extension to data (2010-2013 + FP8)



Salvatore Mele | CERN | Dec. 7, 2009



PARSE.Insight



<http://www.parse-insight.eu/>

- FP7 project, 9 partners, 1.2M€, 3/2008-4/2010:
 - Science (CERN, ESA, STFC, MPG)
 - Libraries (KB, DNB, ...)
 - Publishers (STM)
- Interdisciplinary study on research-data preservation
- Detailed case studies: HEP (+Earth observation, +...)
- Deliverables to inform FP7/8 policy and strategy:
 - Threats/opportunities in data preservation today
 - Roadmap for the future



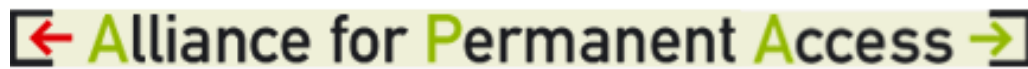
Salvatore Mele | CERN | Dec. 7, 2009



PARSE.Insight survey

Just survey what HEP thinks of preservation...
... but it aims to shed light on:

- Motivations vs. Concerns
 - Threats vs. Opportunities
 - Wishes vs. Obstacles
-
- To provide evidence to decision-making process
 - To provide input to the DPHEP initiative
 - To know what the “community” thinks



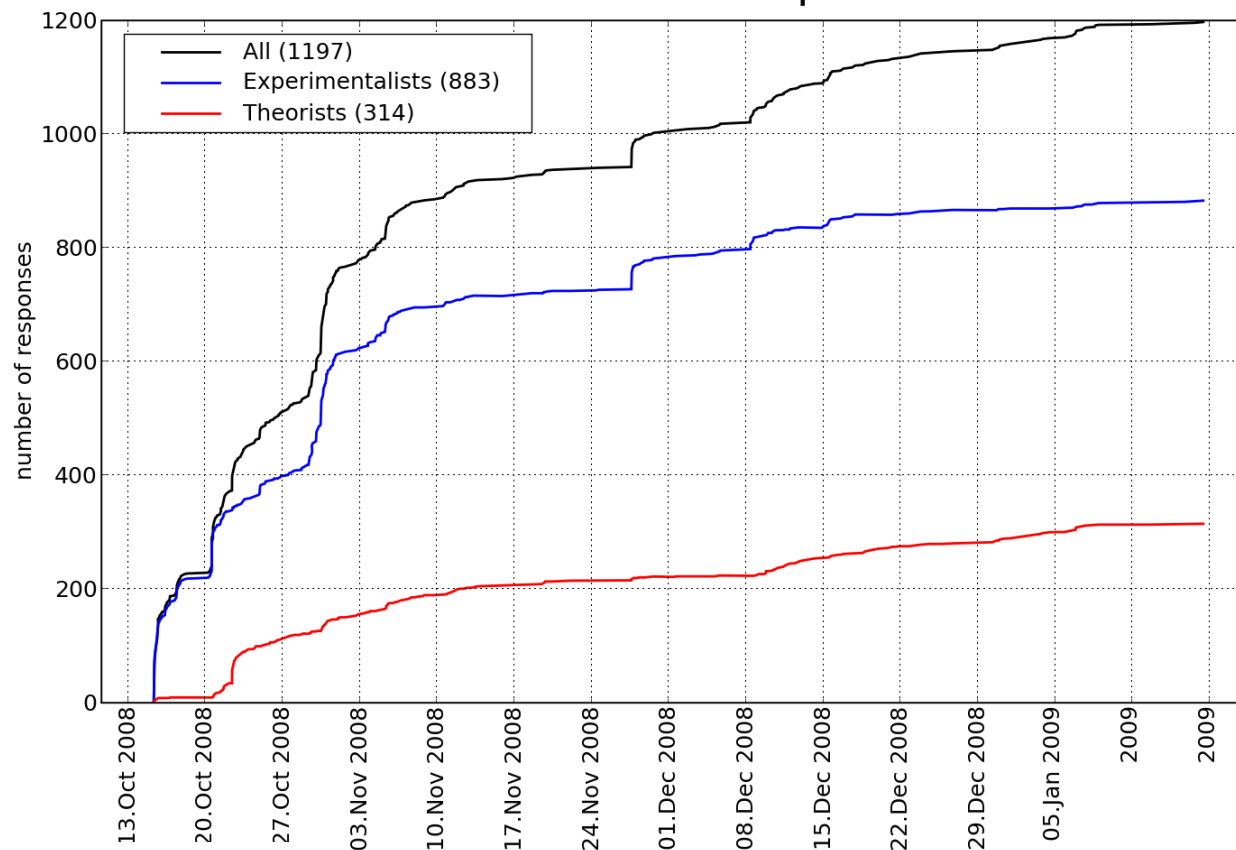
Salvatore Mele | CERN | Dec. 7, 2009



Survey strategy and response

- Livetime of 3 months 10/08-01/09
- Collaborations/theory mailing lists + SPIRES
- 1'200 answers (74% exp., 25% th.). Target ~20'000

Cumulative number of responses vs. time



Survey structure

1. Demographics
2. The importance of preservation
3. What to preserve
4. When, how and where to preserve
5. Threats



Salvatore Mele | CERN | Dec. 7, 2009



Survey structure

1. Demographics
2. The importance of preservation
3. What to preserve
4. When, how and where to preserve
5. Threats

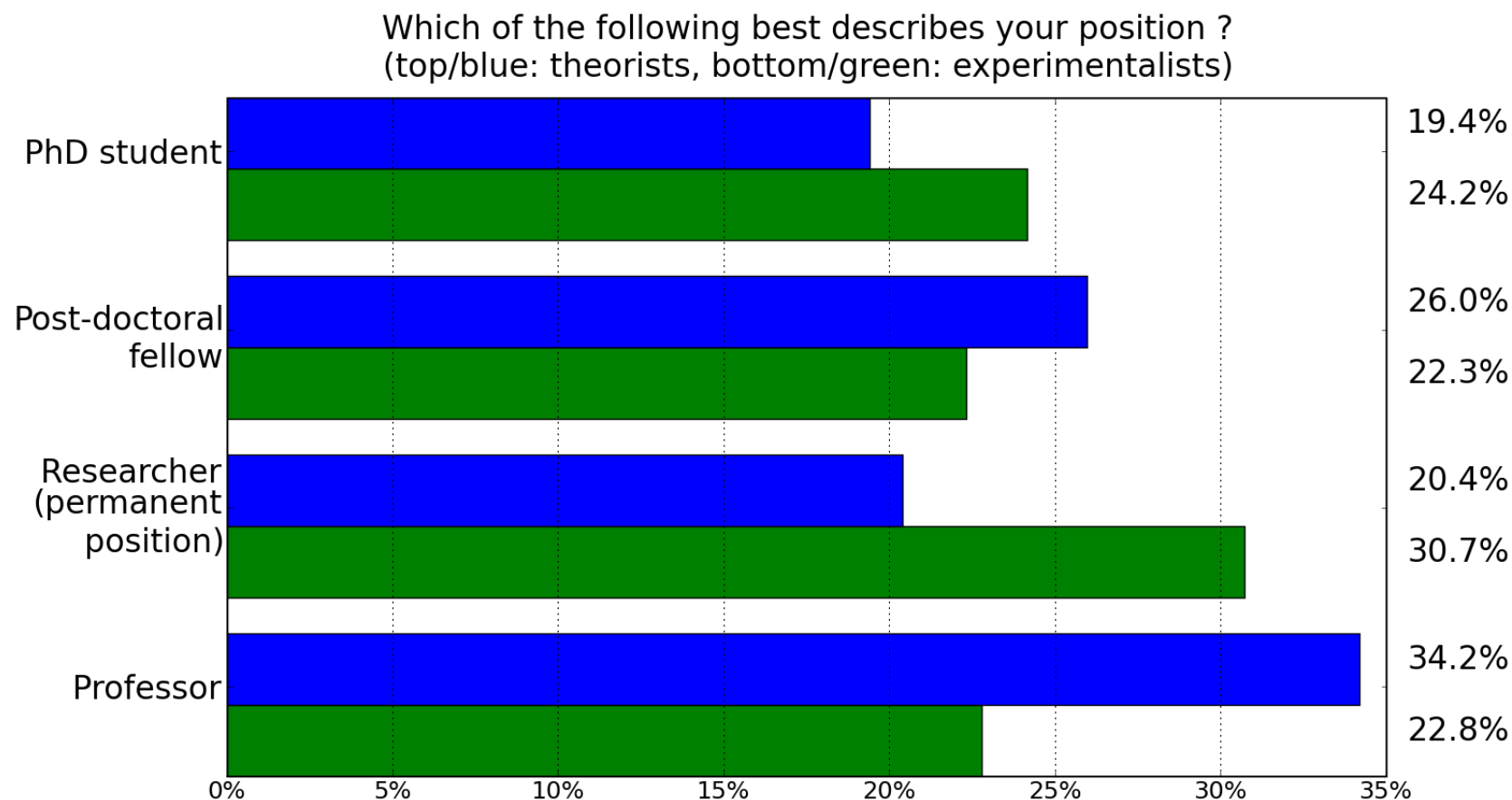


Salvatore Mele | CERN | Dec. 7, 2009



Survey demographics

Reflects the demographics the community



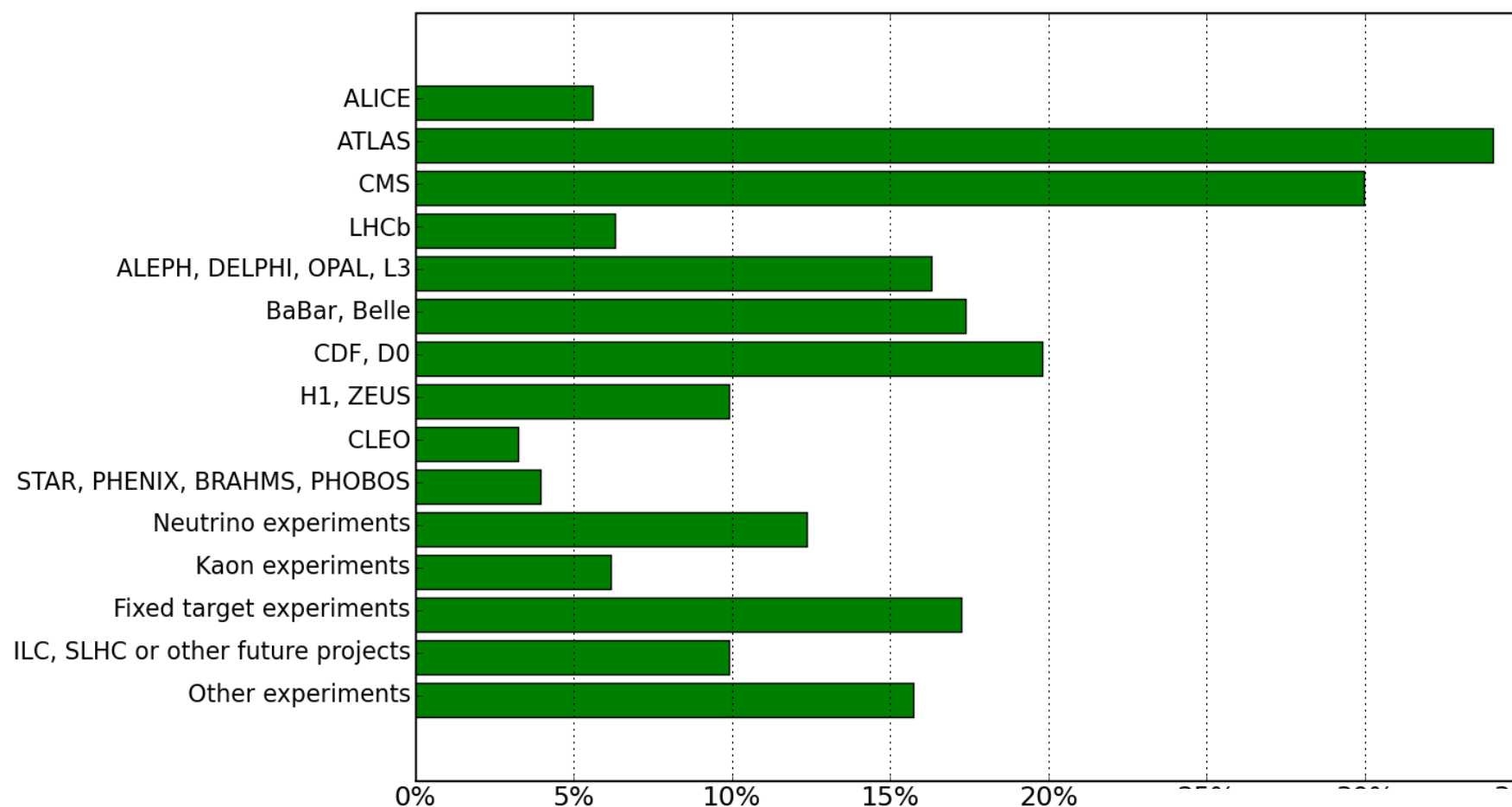
Salvatore Mele | CERN | Dec. 7, 2009



Survey demographics

Reflects the demographics of experiments

In which experiments are you / have you been involved ?



Salvatore Mele | CERN | Dec. 7, 2009



Survey structure

1. Demographics
- 2. The importance of preservation**
3. What to preserve
4. When, how and where to preserve
5. Threats

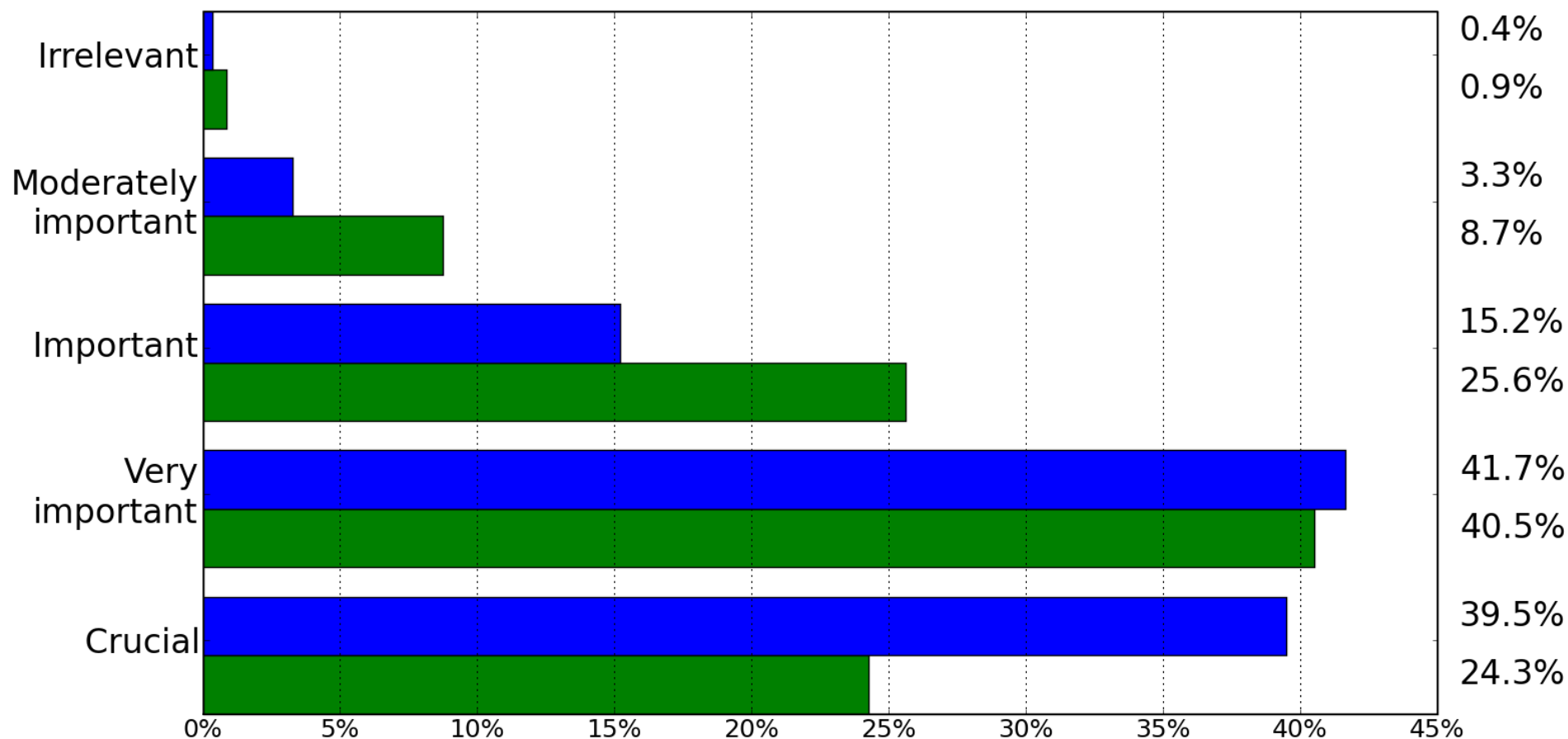


Salvatore Mele | CERN | Dec. 7, 2009



The importance of preservation

In your opinion, how important is the issue of data preservation ?
(top/blue: theorists, bottom/green: experimentalists)

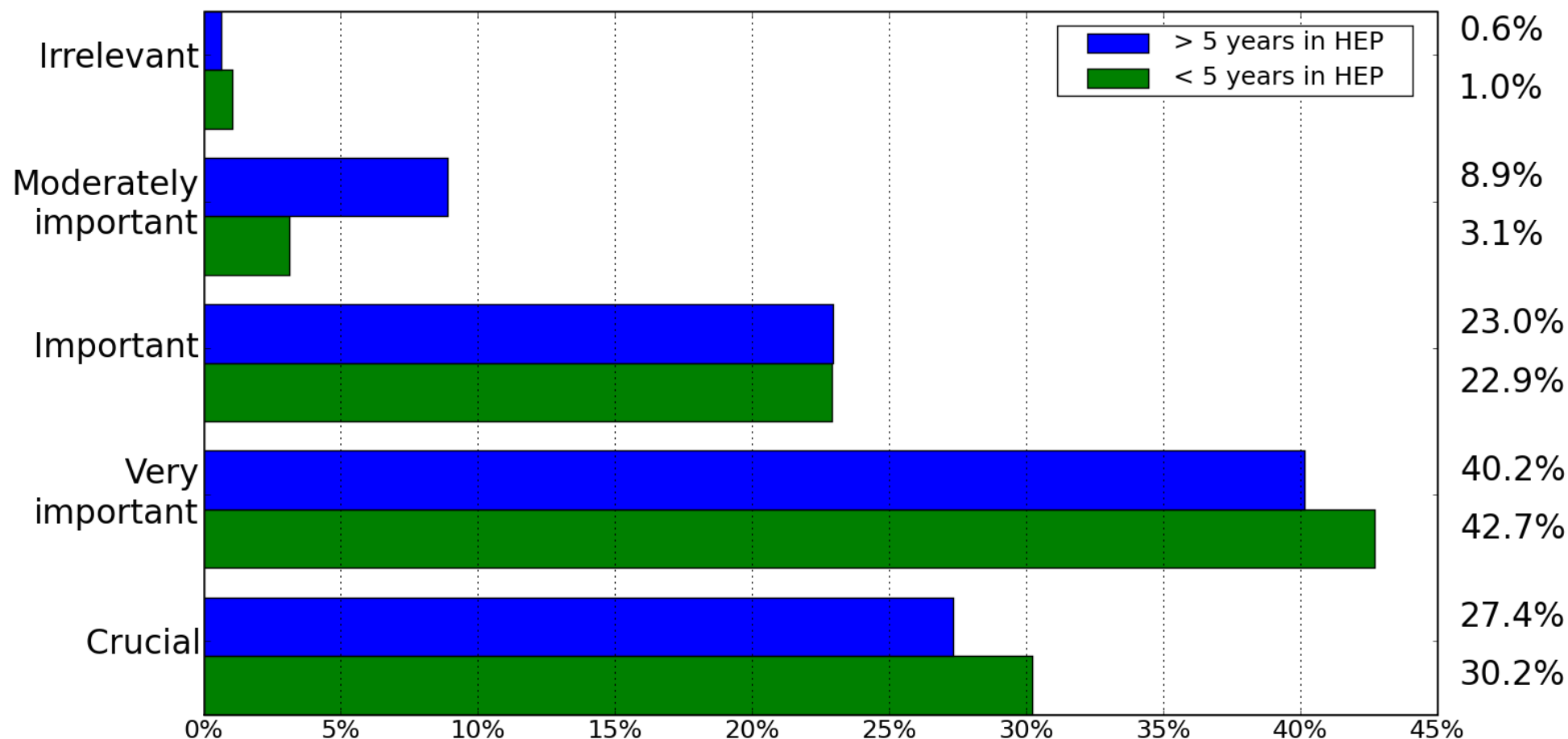


Salvatore Mele | CERN | Dec. 7, 2009



The importance of preservation

In your opinion, how important is the issue of data preservation ?

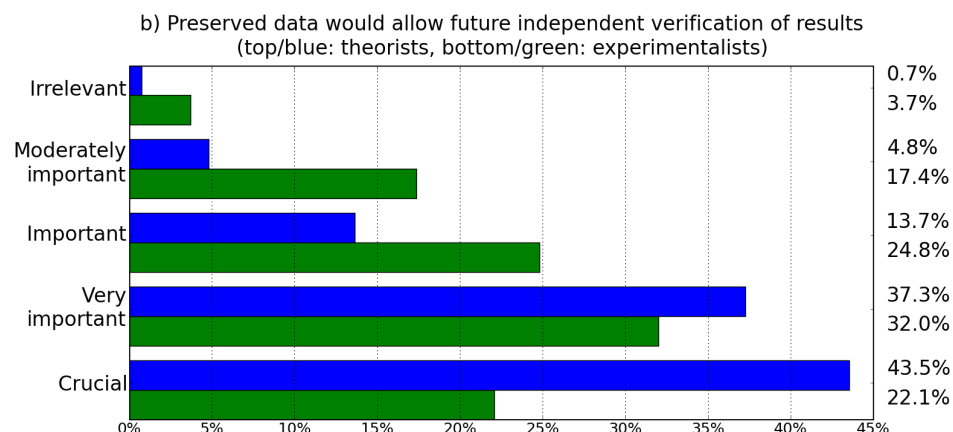


Salvatore Mele | CERN | Dec. 7, 2009

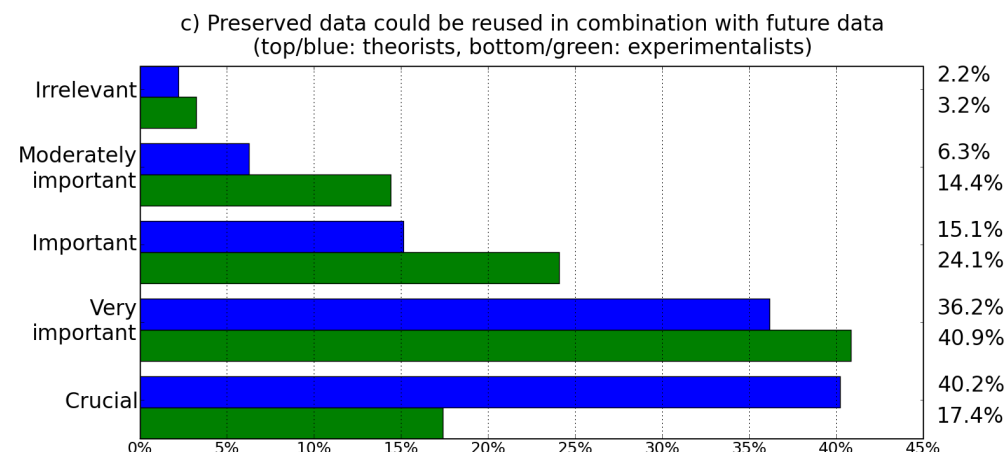


The importance of preservation

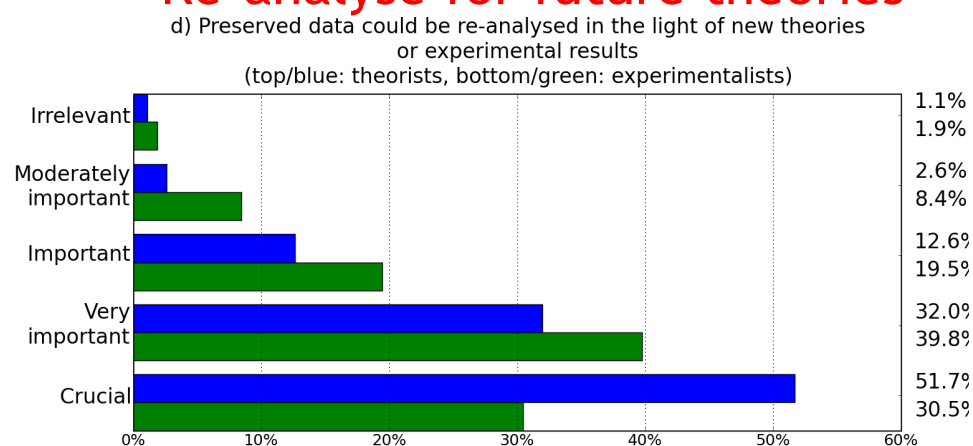
Future independent checks



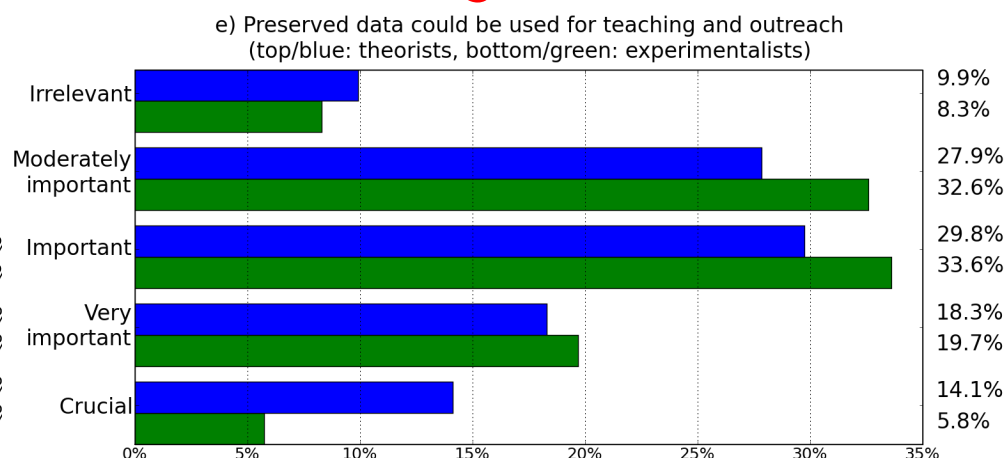
Combine with future data



Re-analyse for future theories



Teaching and outreach



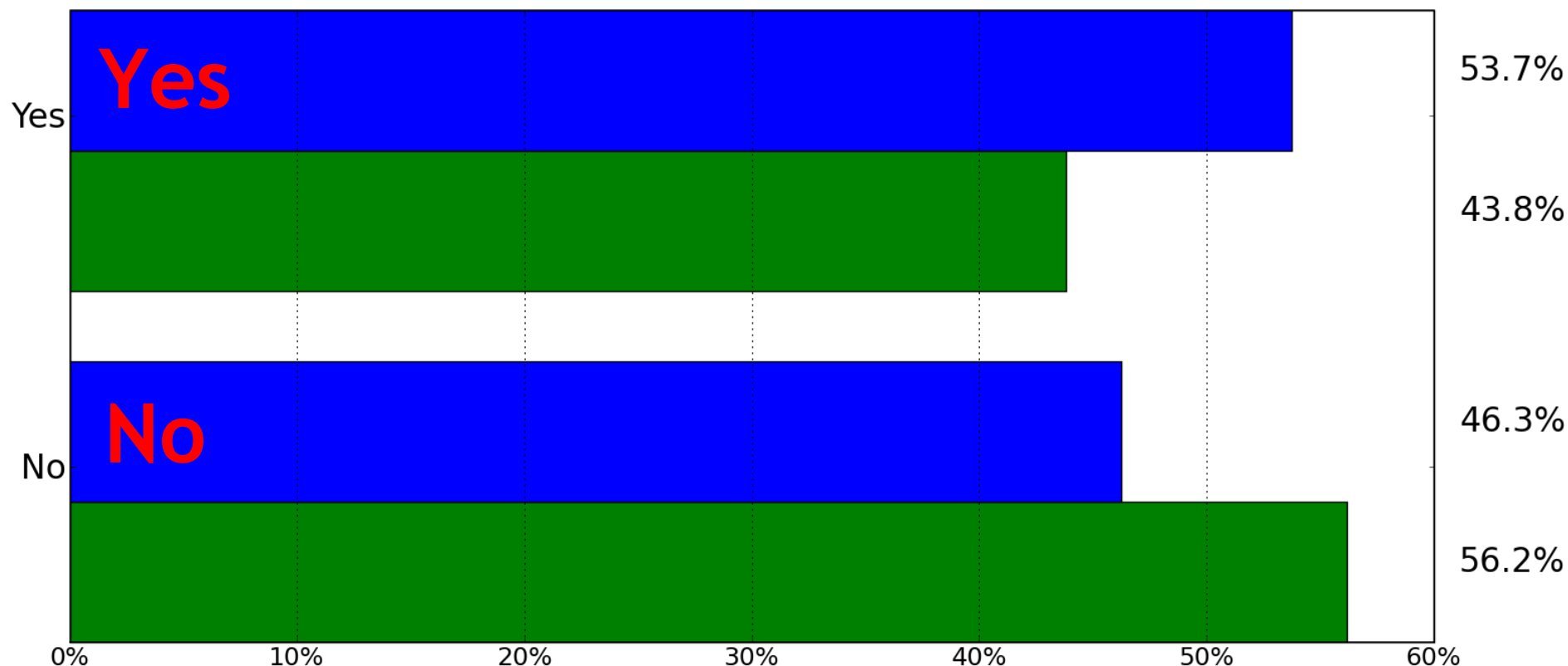
Salvatore Mele | CERN | Dec. 7, 2009



Would preserved data enable better science?

Do you think that access to data from past experiments could have improved your scientific results ?

(top/blue: theorists, bottom/green: experimentalists)



1. Strong argument to ask support for preservation
2. Cannot separate preservation, re-use and access



Salvatore Mele | CERN | Dec. 7, 2009



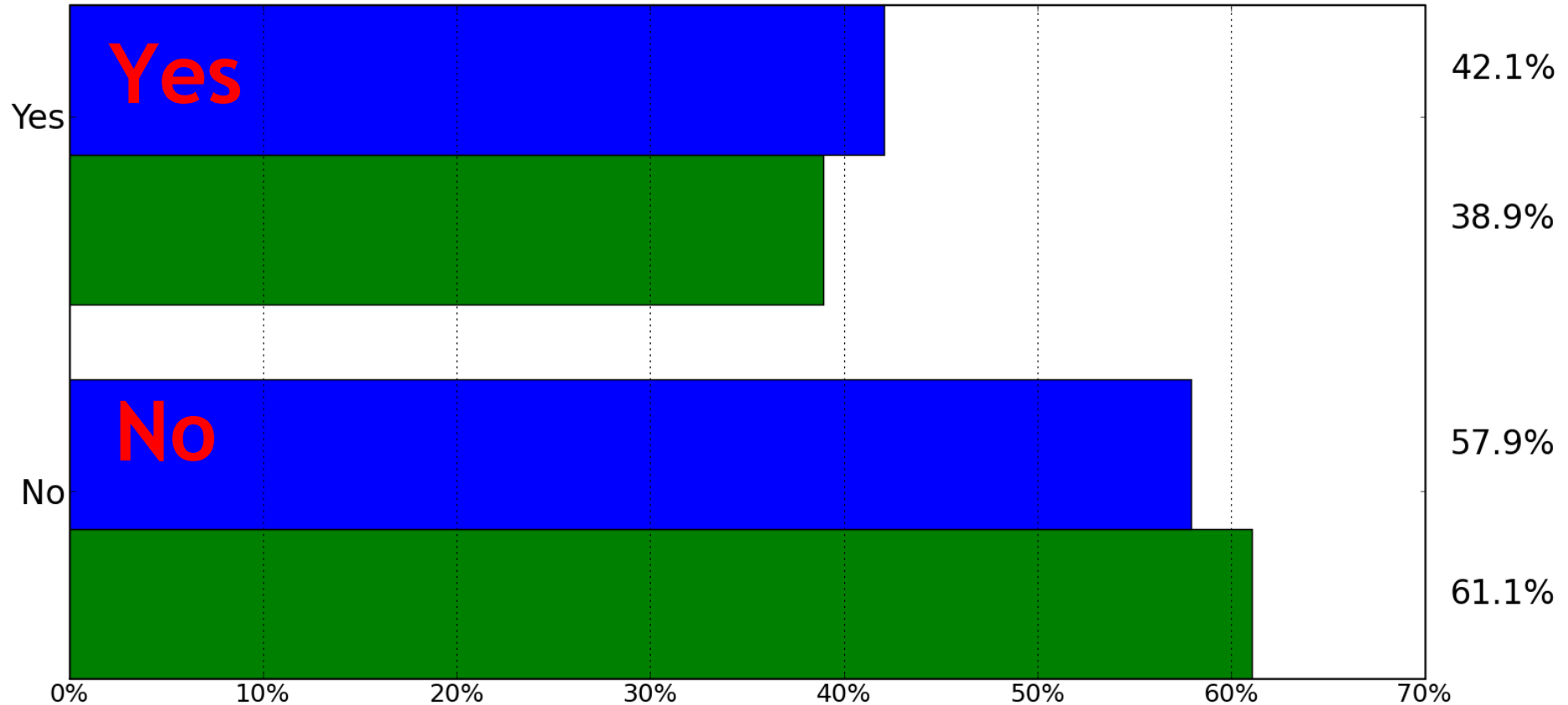


Salvatore Mele | CERN | Dec. 7, 2009



Did anything go wrong so far?

Do you think that in the past important HEP data have been lost ?
(top/blue: theorists, bottom/green: experimentalists)



Over optimistic? Over pessimistic?



Salvatore Mele | CERN | Dec. 7, 2009



Survey structure

1. Demographics
2. The importance of preservation
- 3. What to preserve**
4. When, how and where to preserve
5. Threats

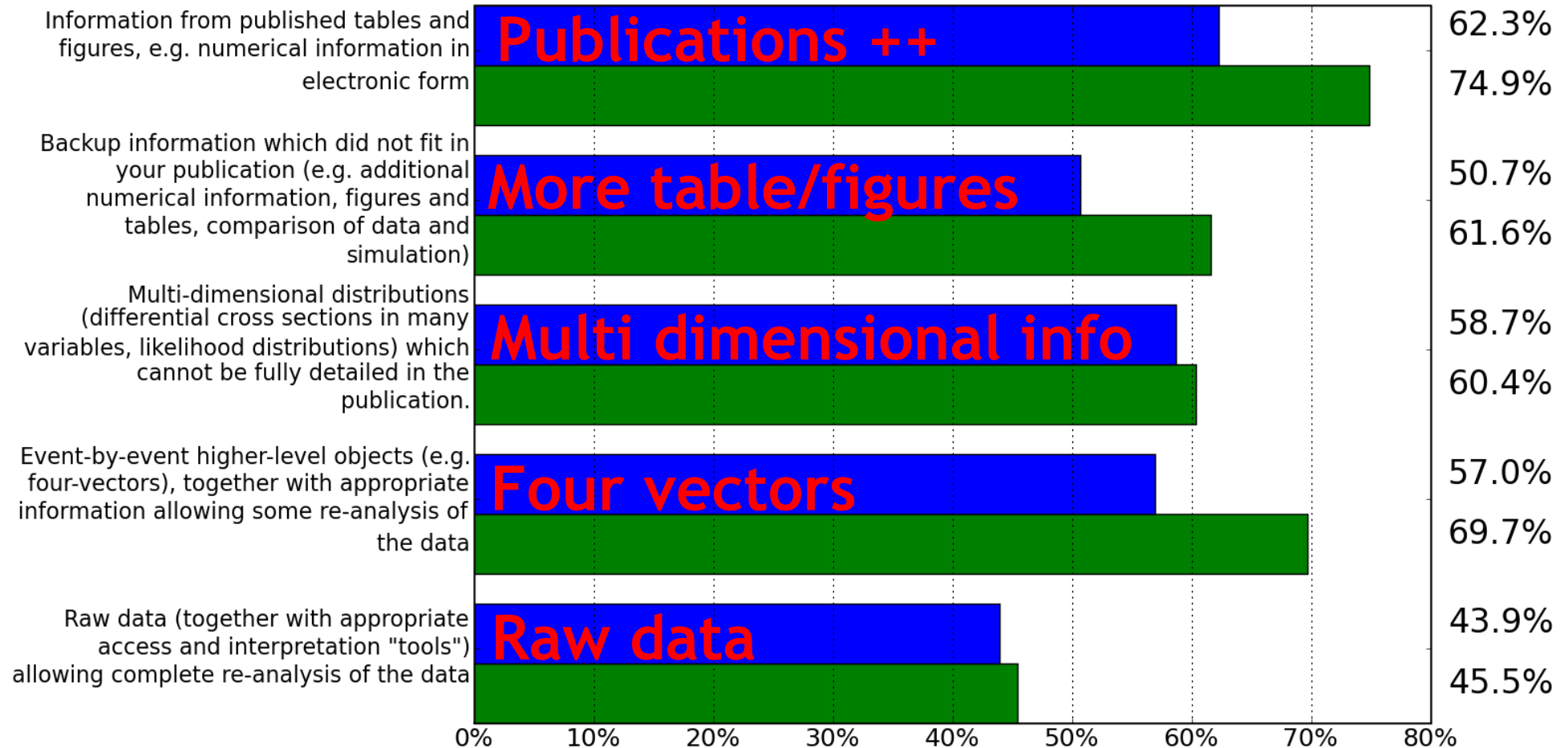


Salvatore Mele | CERN | Dec. 7, 2009



What to preserve?

At what level of detail should data be preserved ?
(top/blue: theorists, bottom/green: experimentalists)



Salvatore Mele | CERN | Dec. 7, 2009



Survey structure

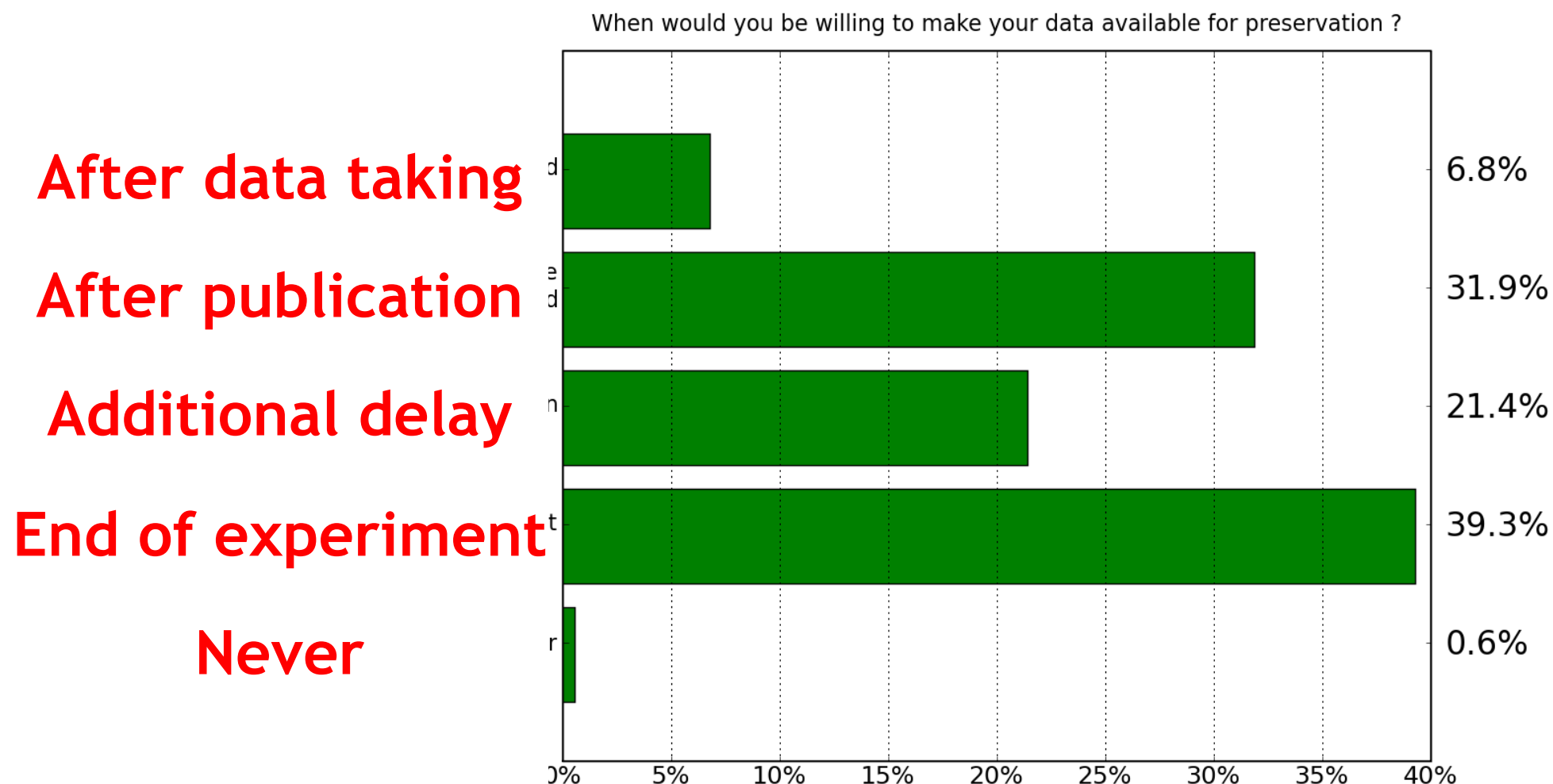
1. Demographics
2. The importance of preservation
3. What to preserve
4. When, how and where to preserve
5. Threats



Salvatore Mele | CERN | Dec. 7, 2009



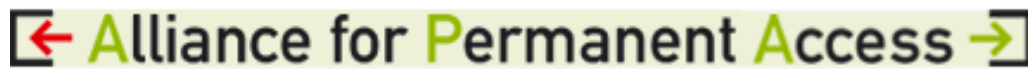
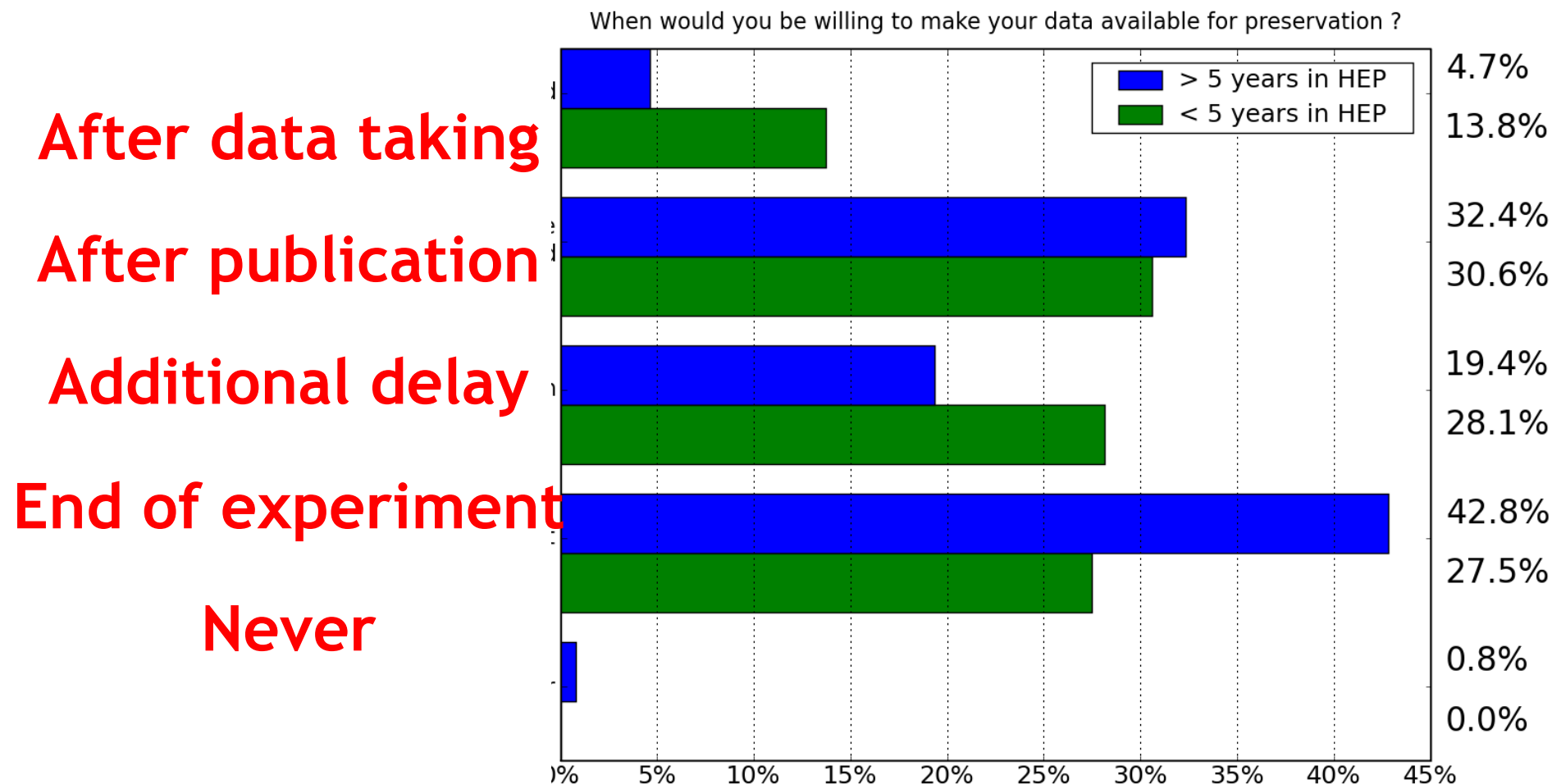
When would data be available for preservation?



Salvatore Mele | CERN | Dec. 7, 2009



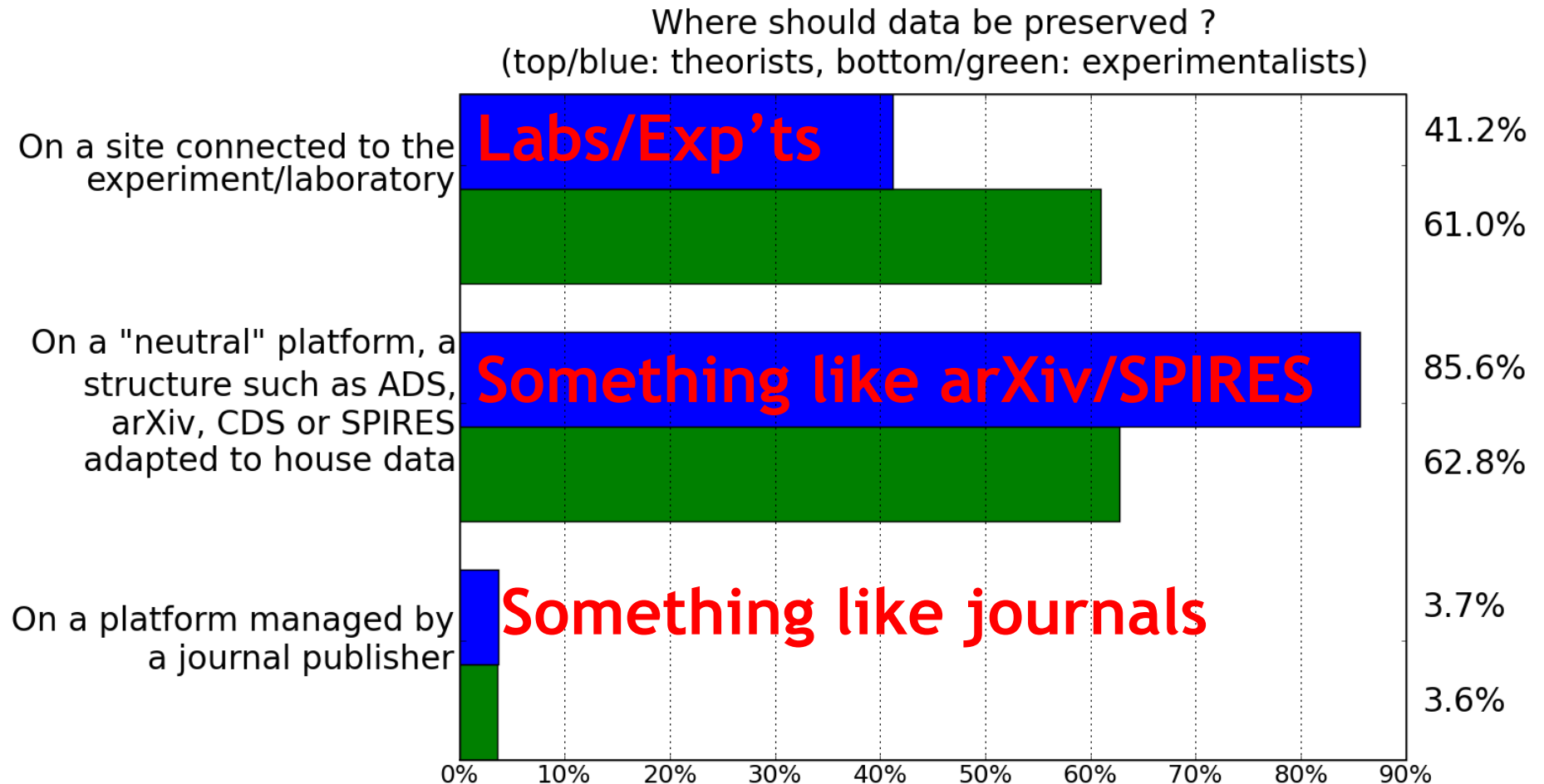
When would data be available for preservation?



Salvatore Mele | CERN | Dec. 7, 2009



Where to preserve?



Survey structure

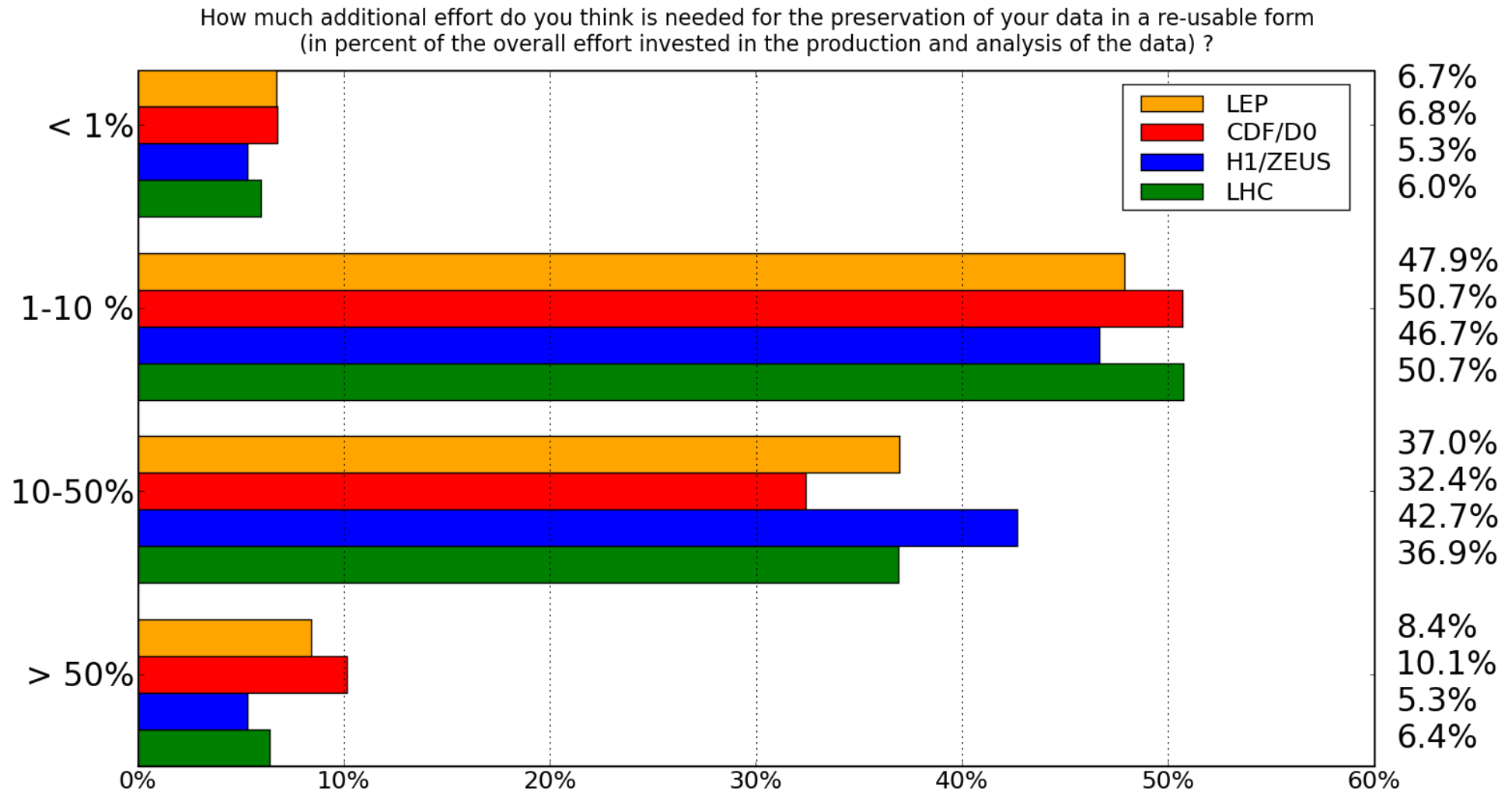
1. Demographics
2. The importance of preservation
3. What to preserve
4. When, how and where to preserve
5. Threats



Salvatore Mele | CERN | Dec. 7, 2009



Reality check #1: how tough is it to preserve?



As a percentage of the data-taking effort

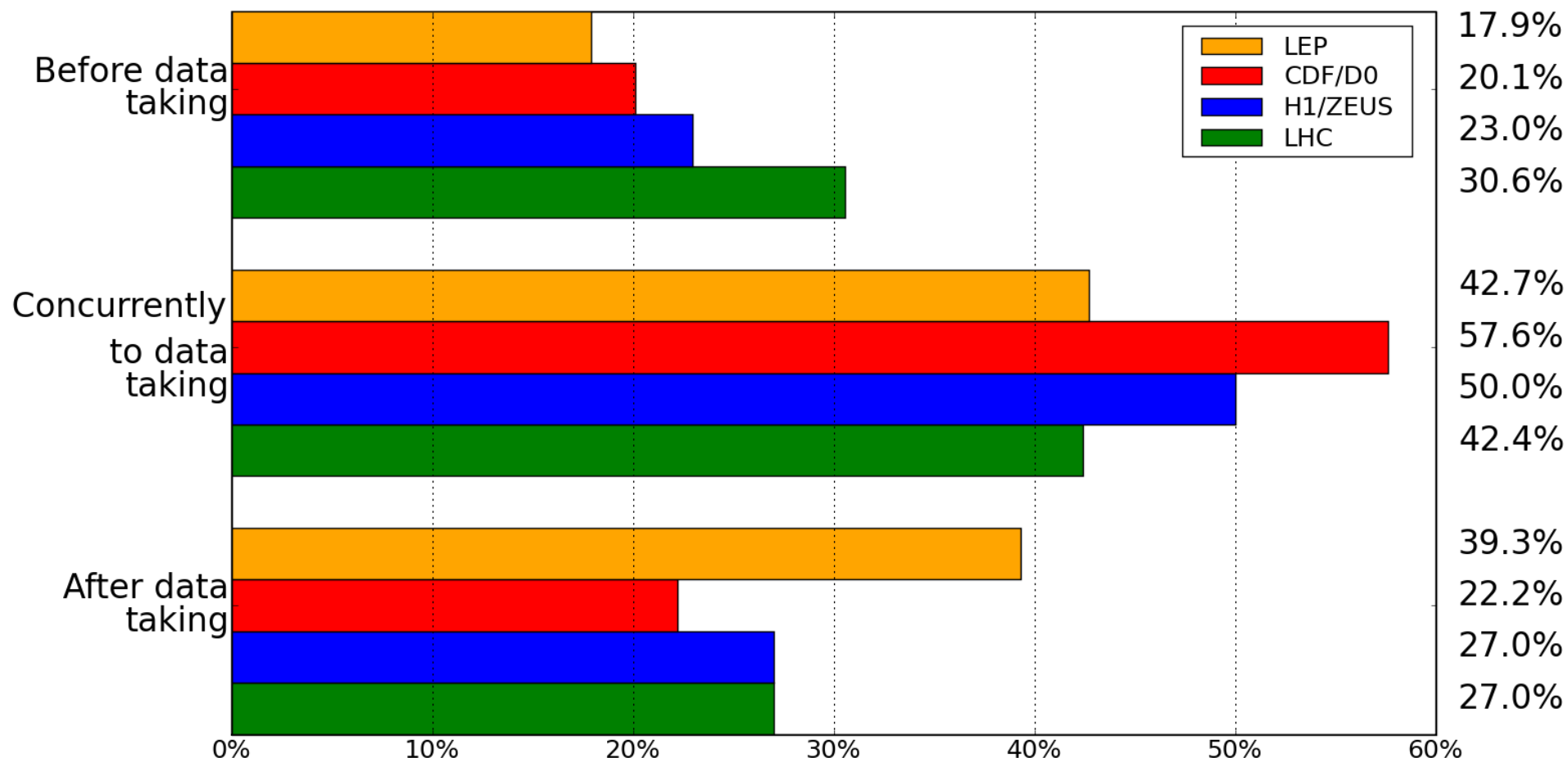


Salvatore Mele | CERN | Dec. 7, 2009



Reality check #2: when to start?

In your opinion, when should this effort start in order to be the most effective ?



....which means now!

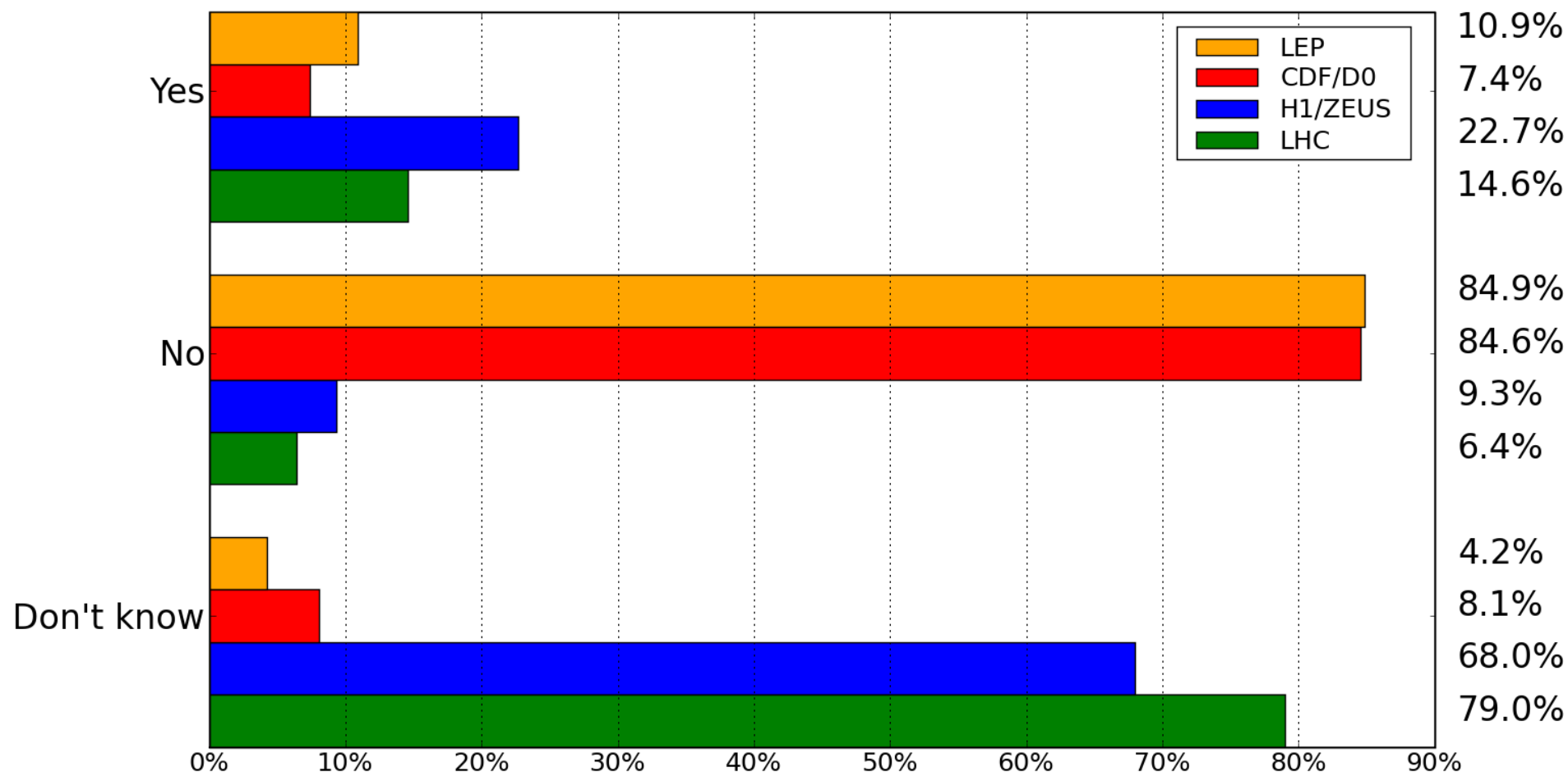


Salvatore Mele | CERN | Dec. 7, 2009



Reality check #3: is it doable?

Will your experiment/collaboration/organisation be able to invest this effort ?



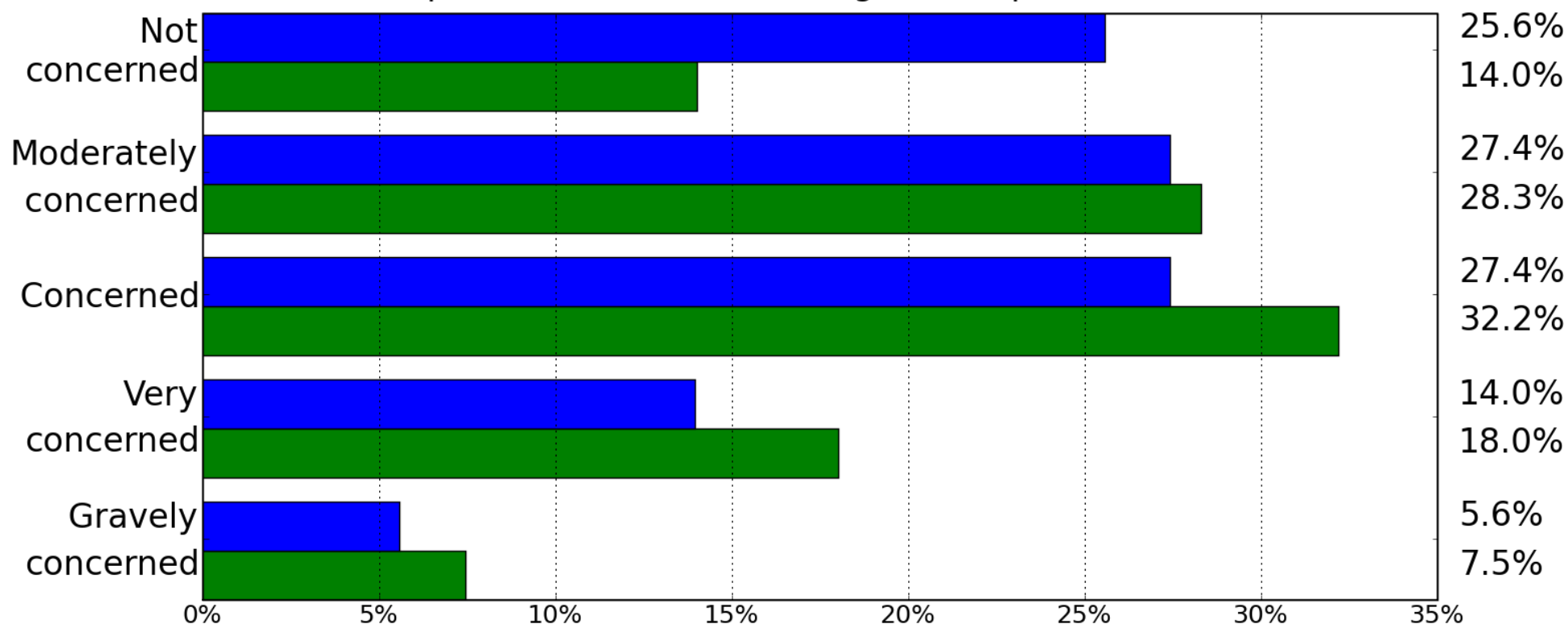
Salvatore Mele | CERN | Dec. 7, 2009



Ideal-case worries: getting credit

To what extent are you concerned about the following issues related to giving access to preserved data ?

a) Preserved data could be used without giving proper credit to the original authors
(top/blue: theorists, bottom/green: experimentalists)



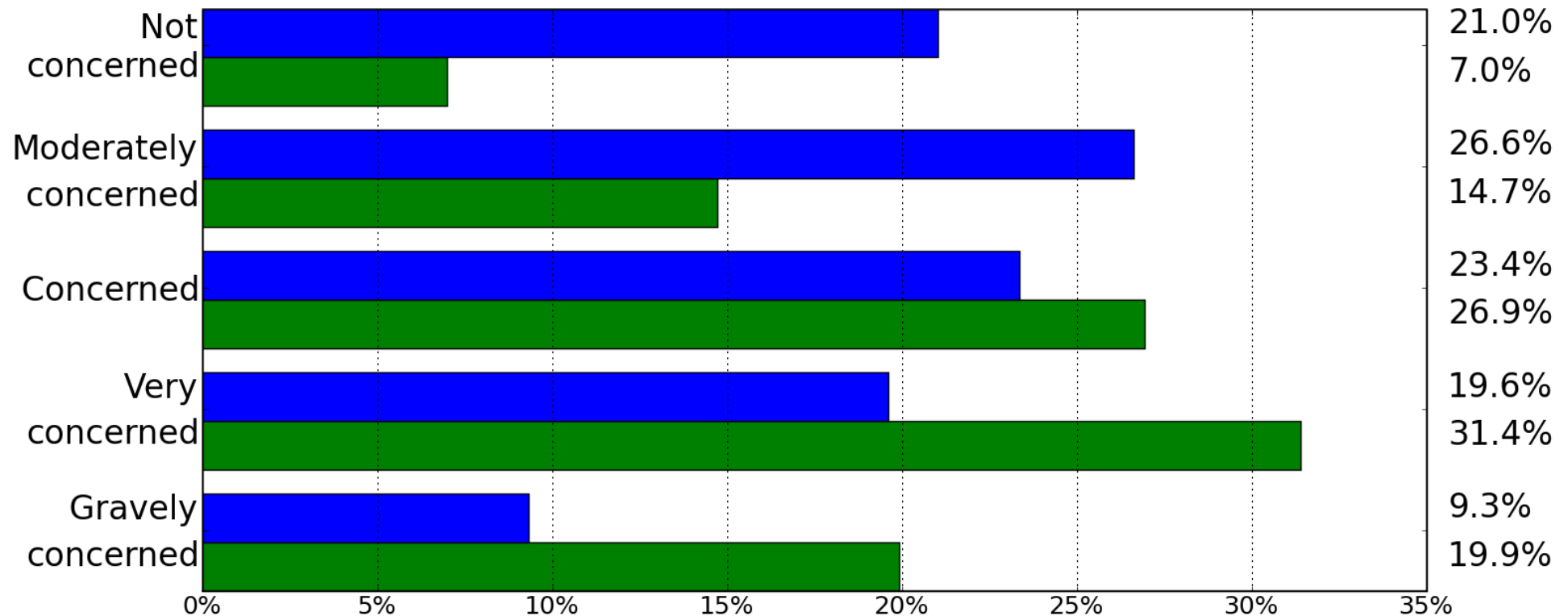
Salvatore Mele | CERN | Dec. 7, 2009



Ideal-case worries: inflation/noise

To what extent are you concerned about the following issues related to giving access to preserved data ?

b) Uncontrolled access to data may lead to an inflation of incorrect results
(top/blue: theorists, bottom/green: experimentalists)



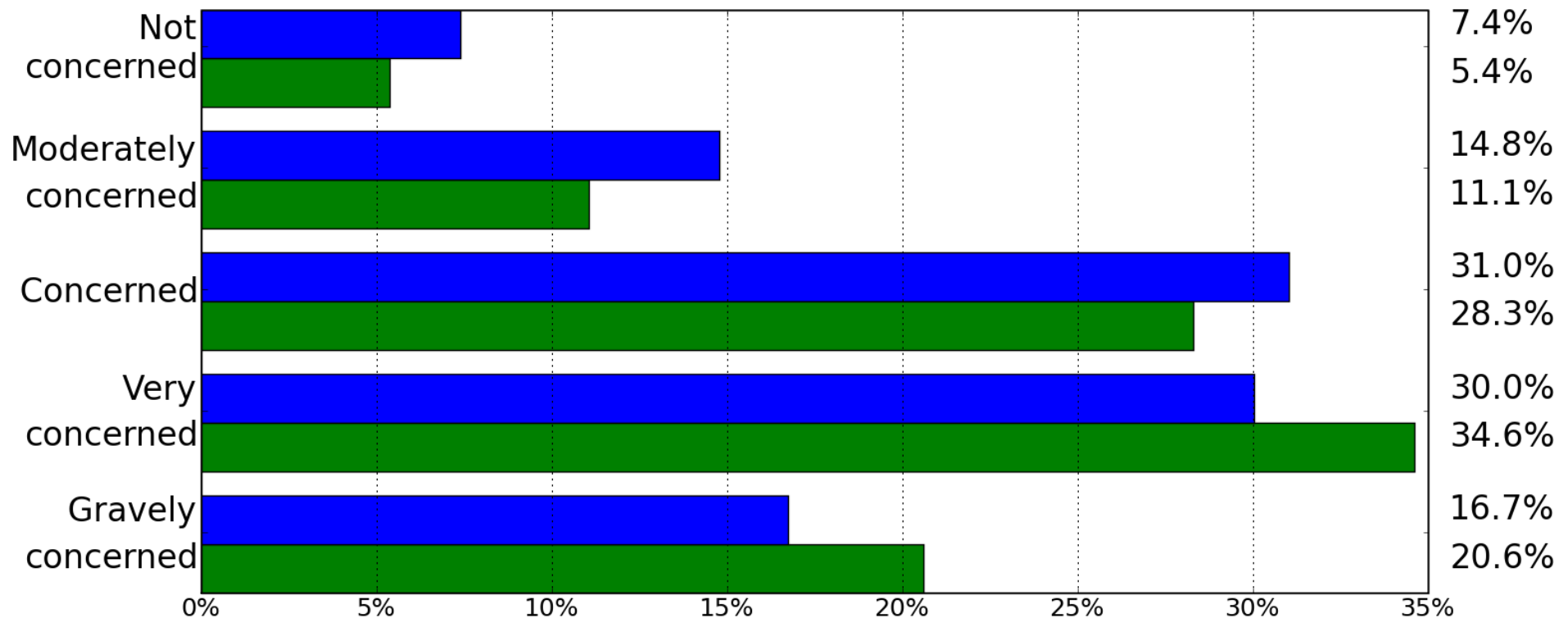
Salvatore Mele | CERN | Dec. 7, 2009



Ideal-case worries: documentation

If you were to re-use preserved data, to what extent would you be concerned by the following scenarios ?

d) I am not using the data correctly
(top/blue: theorists, bottom/green: experimentalists)



Salvatore Mele | CERN | Dec. 7, 2009



Conclusions

Provided quantitative insight on needs and opportunities in data preservation in HEP

- **Strong awareness/support of necessity/urgency of data preservation**
- **Consensus across experiments and with theory**
- **Suggestion that global initiatives can be the solution**



Salvatore Mele | CERN | Dec. 7, 2009



Thank you!

Salvatore.Mele@cern.ch

**More details on PARSE.Insight HEP survey:
<http://arxiv.org/abs/0906.0485>**



Salvatore Mele | CERN | Dec. 7, 2009

