TASK 3- DATA COLLECTION METHODOLOGIES

- The aim of this task is to propose up to 3 alternative methods for collecting the information identified under Task 2.
- To be able to compare the different methodologies a table will be provided showing for each of them their advantages/weaknesses.
- Cost estimates for each alternative method will be provided accompanied by a justified estimation on how much the cost may vary among Member States.
- Existing harmonised statistics on road accidents in EU will be studied and compared to the proposed methodologies.
- This task also includes defining the methodologies to complete missing data. This includes considerations related to the reliability and uncertainty of the data collected.

TASK 3 - TEAM

TASK 3 - CONTEXT FOR COLLECTION OF FIRE DATA

- Why collect fire data?
  - Information about fires (how, when, where, how) and at-risk populations is valuable for intervention, education, budgets, training, etc.
  - Data needs to accurately identify the magnitude and characteristics of the fire problem to be useful.

- Important considerations:
  - Fire departments are principal data providers. Data is provided by firefighters who have other demands on time.
  - Experience with fire data collection and ability to support national data collection system varies among EU Member States.

TASK 3 - DATA COLLECTION METHODS: CENSUS SURVEY

- A census is a survey conducted on the full set of observation subjects belonging to a given population or universe.
- Result
  - This method, if successful, will result in a count of total number cases (e.g., all fires, all deaths, all injuries, etc.).
  - Provides a full database of all incidents in addition to total case numbers.
- Advantages:
  - A true census will ideally capture all cases, though details for each case might be incomplete.
  - Avoids risk of sampling errors
  - Provides data for small areas (critical for planning services)
- Disadvantages:
  - Size and complexity of design and implementation - Total coverage of the population is required for a true census.
  - High cost

TASK 3 - DATA COLLECTION METHODS: SAMPLE SURVEY

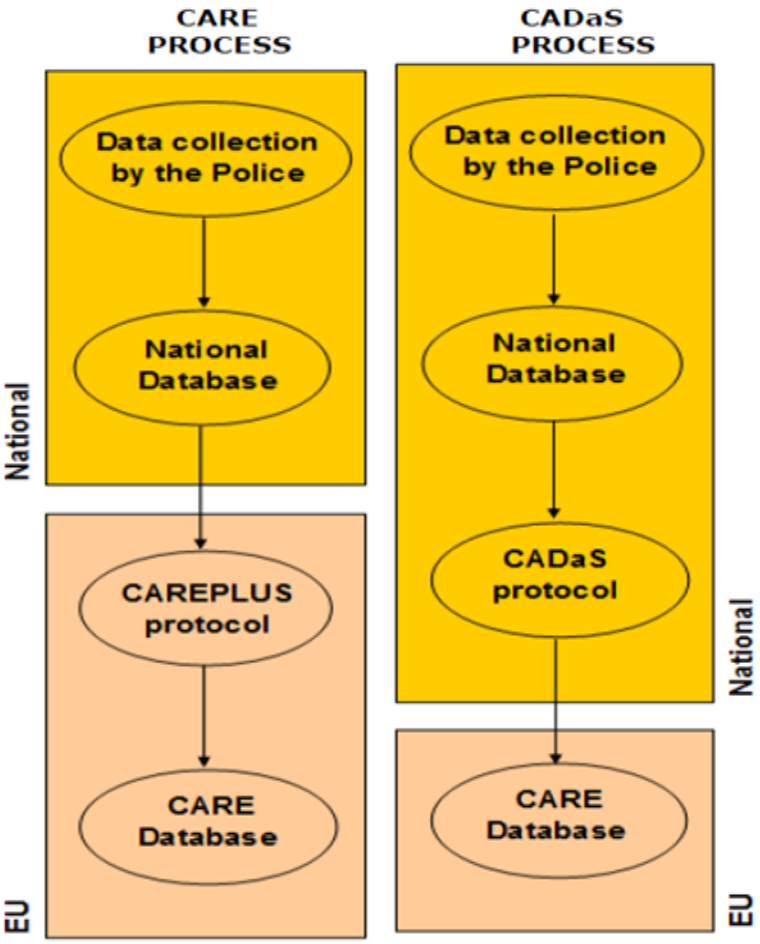
- A sample survey is a survey which is carried out using a sampling method, i.e. In which a portion only and not the whole population is surveyed.
- Result:
  - A properly designed sample will result in an estimate for the total number of cases (e.g., estimated total fires, estimated total deaths, estimated injuries, etc.).
- Advantages:
  - Less costly than census
  - Totals can be compared to totals from Census
- Disadvantages:
  - Relative to census, data on sub-populations or small geographical areas may be unreliable
  - Estimates are subject to sampling error
  - Proper statistical sample design is required.
  - The sample must be representative of the population for whom the estimate will apply (e.g., cannot sample within urban areas and expect the estimate to apply to rural communities).
  - Inclusion or omission of serious incidents can skew the loss measures.
  - Does not include full database of incidents.

TASK 3 - DATA COLLECTION METHODS: CONVENIENCE SAMPLE

- A convenience sample is a form of non-probability sample - selection of units based on factors other than random chance. Convenience sampling relies on respondents who are convenient to researcher.
- Result
  - This method will NOT result in a total (estimated or not).
  - Descriptive statistics for the collected data can be reported, but those results cannot be considered representative for the entire population.
  - The full database of what was collected will be available.
- Advantages:
  - Least costly
  - Less complex and easier to implement
  - Data can be collected quickly
- Disadvantages:
  - Results will be biased relative to underlying population
  - National estimate for totals cannot be calculated
  - Completeness of the dataset cannot be determined
  - Likely to exclude large parts of population
  - Not possible to compare result to that of a Census or Sample Survey

TASK3 – EU ROAD ACCIDENT DATA

- The Community database on Accidents on the Roads in Europe (CARE) - detailed data on individual accidents collected by the Member States of the European Union
- Includes information recorded from all road accidents involving at least one moving vehicle and one injury or death
- The quality and availability of road accident data are affected by differences in existing national data systems
- Common Accident Data Set (CADaS) have been developed to compare road accident data
- CADaS contains a minimum set of standardised data elements that can be implemented voluntarily.
- The EU Member States are not obliged to adopt CADaS and they transmit the data at the EU level choosing the level of detail



TASK 3 – COMPARING EU FIRE AND ROAD ACCIDENT DATA

- Both rely on voluntary reporting by emergency responders (Police\Firefighters)
  - Unable to identify any published assessment of the completeness of data set
- Experience with preparing incident report differs
  - Preparing reports of vehicle accidents resulting in casualties is a more traditional practice for police officers and is likely to be an expected part of job responsibilities
  - For firefighters, filling out paperwork on fire incidents is a comparatively novel part of the job, one with less tradition behind it
- Threshold for data collected differs
  - Data collection in CARE limited to road accidents resulting in death or injury
  - Fire incident data collection generally includes all fires attended by fire departments, even those resulting in minimal damage.
- Conclusion
  - CARE appears to be a census for fatal road accidents.
  - This narrower criteria for in-scope incidents in CARE data collection appears to result in more complete reporting.
  - Fire incidents using broader inclusion criteria have higher risk of non-reporting.
  - If it is the goal to have complete reporting of fire incidents in the European Union, consideration should be given to limiting data collection to fatal fires!

TASK 3 - MISSING DATA IN SURVEY RESEARCH

- Missing data are the norm in survey research.

- Missing data take 2 primary forms:
  - *Unit level* missing data is the failure to obtain <u>any</u> information from a respondent.
  - *Item level* missing data refers to the failure of respondents to complete all items in a survey, i.e. individual questions/items not answered.
  - Missing data at the unit level is considered the most problematic because the researcher receives no information that can be used to improve estimates or reduce bias arising from missing data.

- Why is missing data important?
  - Missing data can: reduce statistical power of a study, produce biased estimates, lead to invalid conclusions

- Example: What if missing data on fire incidents comes from rural areas with populations that may be older or structures that have special fire risks?

TASK 3 - MISSING DATA PATTERNS

- Patterns of missingness influence potential for bias:
  - **Missing completely at random:** Missing data unrelated to variable of interest and subjects with complete data indistinguishable from those with incomplete data.
  - **Missing at random:** Missing data partly dependent on other data, but not dependent on missing values. Respondents with incomplete data differ but pattern of missing data can be predicted from other data in dataset.
  - **Missing not at random:** Missing values related to reason that are missing (high income earners likely to ignore question on income). “Non-ignorable” missing data.
- Strategies for dealing with missing data:
  - Try to obtain missing data.
  - Leave out incomplete cases.
  - Replace missing data with conservative estimate (replace missing value with group mean).
  - Estimate from other data: replace missing values with values derived from respondents that matched on other variables or use other statistical techniques (multiple imputation, maximum likelihood).
- Researchers should minimize missing data in advance through careful design of surveys and data collection strategies.

MISSING DATA IN FIRE DATA COLLECTION

United Kingdom

- Tables are footnoted when significant number of incidents determined to be missing.
- Missing data in data fields said to be low and are reported as unknown. No imputation or other estimation methods.
- In Wales, statisticians may impute data recorded as missing.

Sweden

- Assumed no underreporting of fire incidents now occurs.
- Swedish Civil Contingencies Agency (MSB) sends feedback to fire departments with incomplete reports.
- Logic checks prevent reports being sent with missing values.
- Don't use estimation methods to mitigate missing data.

France

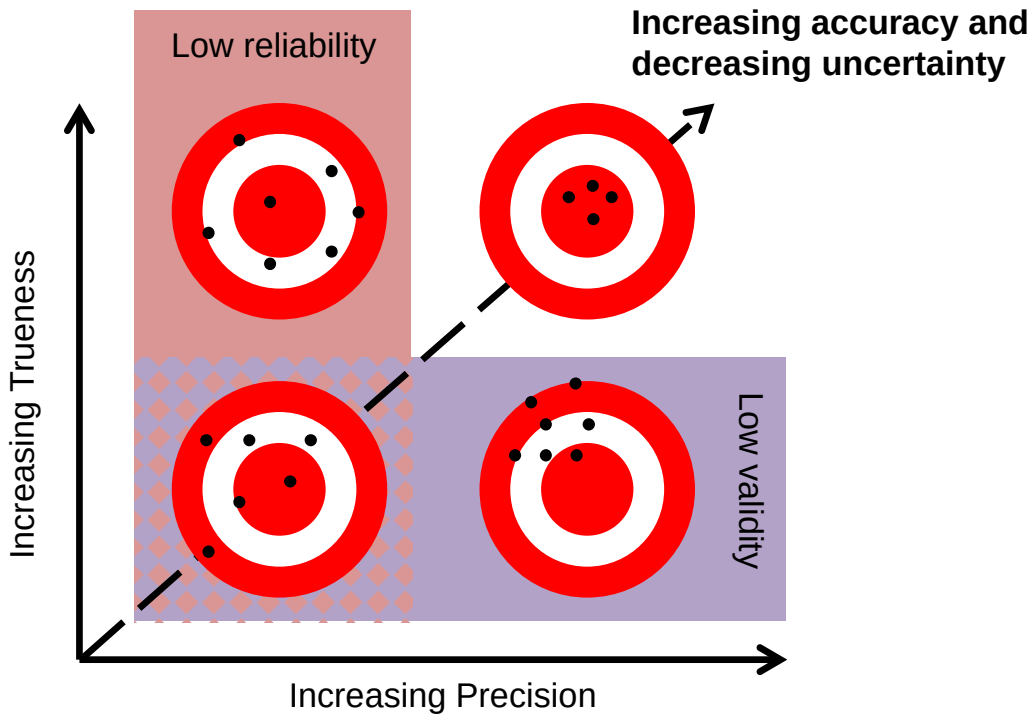
- In cases where fire interventions not reported, Ministry of Interior calculates the weighted average of interventions by fire departments and applies the number for non-reporting departments

United States

- National Fire Incident Reporting System (NFIRS) is a detailed sample of fire incidents from a universe of incidents whose size can't be measured from NFIRS alone.
- A 2nd survey is administered by the National Fire Protection Association to create an estimate of the true size of the universe and scale up NFIRS results.
- For item-level missing or unknown data NFPA allocates unknown or missing data proportionally among known data by calculating percentages based on the known data and then multiplying the totals by the percentages known. This approach assumes that if the missing data were known, it would be distributed in the same manner as the known data.

UNCERTAINTY ANALYSIS AND STATISTICAL RELIABILITY

- Uncertainty relates to how an estimate might differ from a “true value”; a reduced uncertainty will result in an increased accuracy, where accuracy is defined as the closeness of agreement between a measured value and the true value



UNCERTAINTY ANALYSIS AND STATISTICAL RELIABILITY

- Different type of errors lead to increased uncertainty
  - Sample error
  - Non-sample error
- Non-sampling errors are usually very difficult to quantify, and specific studies are usually required to make such assessments. Examples of non-sampling errors include:
  - Cases when data is not collected
  - Faulty inputs due to ignorant reporter, or fatigue when reporting
  - Inaccurate answers (forms filled in inaccurately, questions misinterpreted, definitions misunderstood)
  - Errors when processing, presenting, or analysing data

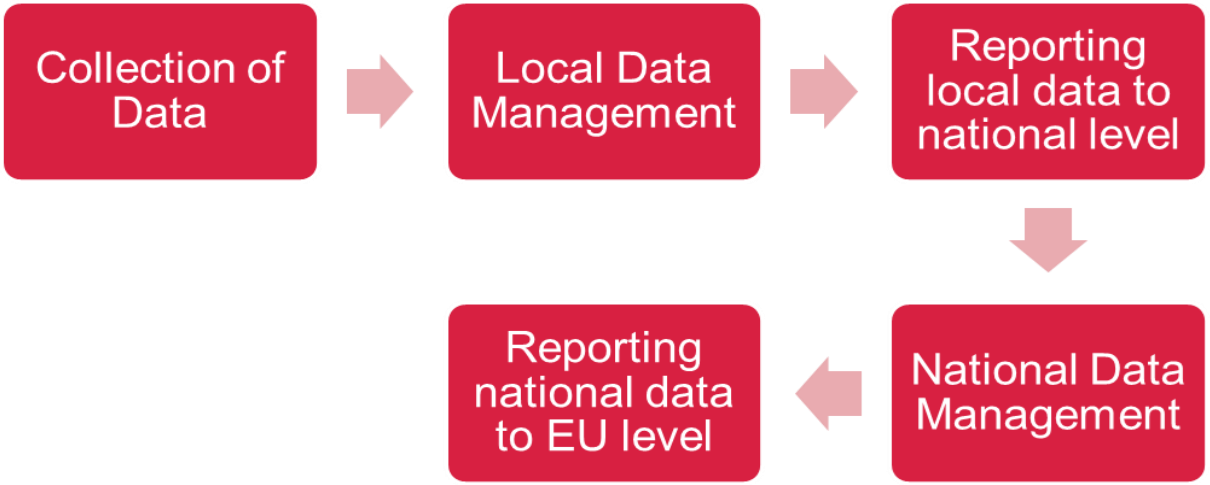
Data collection method	Sampling errors	Non-sampling errors
Census	Limited	Yes
Sample survey	Yes, but can be treated	Yes
Convenience collection	Yes	Yes

UNCERTAINTY ANALYSIS AND STATISTICAL RELIABILITY

- Based on the discussion in the report some some general estimates of the uncertainty associated with the different variables suggested in Task 2.
- These are rough qualitative estimates that indicate which variables can be expected to be affected by the largest degree of uncertainty.
- It should be stressed that these uncertainties can be reduced by applying different measures, like the measures discussed in the report.

Variable	Estimated associated uncertainty
Number of fatalities	**Medium**
Number of injuries	**High**
Fire cause	**High**
Type of building	**Medium**
Incident location	**Low**
Incident date	**Low**
Incident time	**Low**
Age of victim	**High**
Number of floors	**Low**
Room of origin	**Low**
Source of ignition (or heat source)	**High**
Material mainly responsible for fire development	**Medium**

TASK 3 – BASIC PREMISE FOR COST CALCULATIONS



Regardless of the data collection methodology proposed, reports must be collected at the response level for each incident.
The major cost difference between the 3 methodologies is related to what data are reported and how that data are managed at national level.
Focus of cost calculations are therefore on the national level!

TASK 3 – CALCULATION OF COST OF CENSUS METHOD AT NATIONAL LEVEL

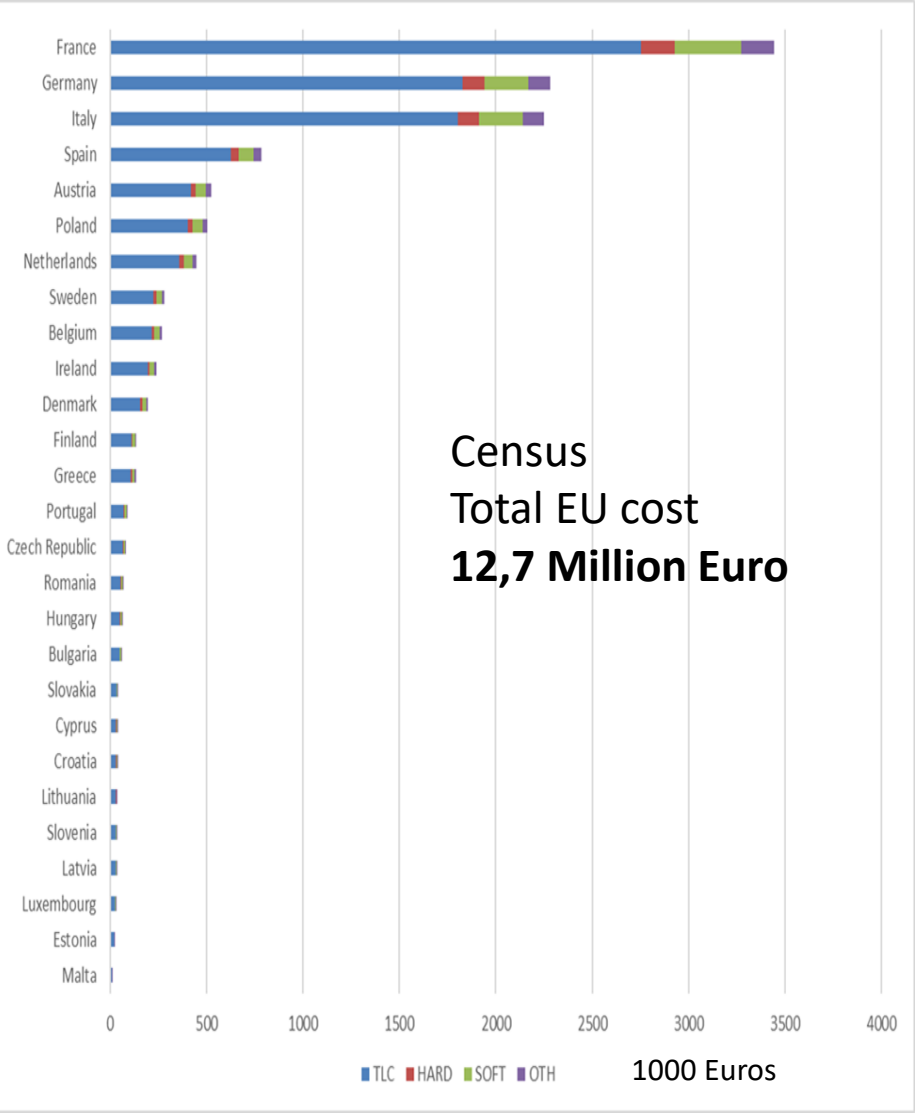
- Information needed to calculate cost:
  - Number of fires (events) [FIRE]
  - Average annual hours worked per worker (hours per year) [AHW]
  - Labour cost from Eurostat (€ per hour) [LCH]
  - Average hours worked per fire event (hours per event) [AHFE]
  - AHFE estimated to **0.25 hours/event** based on CTIF experience
- Cost Factors Calculated:
  - Total hours used for yearly fire statistics THP = FIRE x AHFE
  - Total labour costs TLC = THP x LCH
- Based on CTIF surveys the following ratios are used to estimate the total costs:
  - [TLC] - the total labour costs (in €) are included in the calculation with 80%.
  - [HARD] – Computer Hardware is included in the calculation with 5%.
  - [SOFT] - Software is included in the calculation with 10%.
  - [OTH] - Other costs, including training, refresher courses, etc. 5%
- The total costs of the national fire statistics are ultimately determined from the following relation:
  - **TOTAL = TLC + HARD + SOFT + OTH**

TASK 3 – CALCULATION OF COST OF SURVEY SAMPLE METHOD AT NATIONAL LEVEL

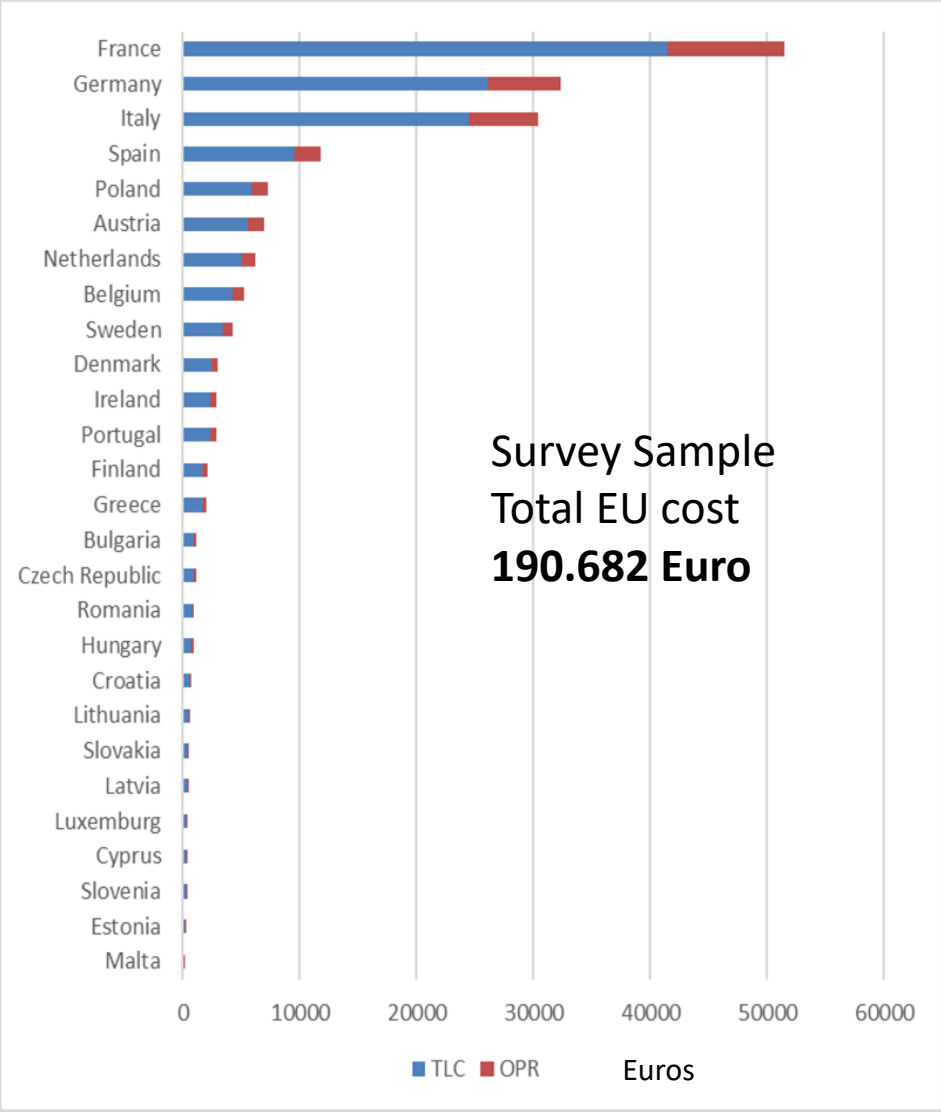
- Using the data from NFPA's Fire Experience Survey, the parameters needed to calculate cost in each EU member state was found to be:
  - Average hours needed per fire event [AHFE] = **0.0035 hours/event** – this is scaled to all fires not just sample size!
  - The operational cost [OPR] amounts to approximately 24% of the labor cost.

- Using these values, the cost at the national level for a survey sample collection of fire data per country can be calculated as follows:
  - Total hours for performance of yearly fire survey sample THP = AHFE x FIRE
  - Total Labour Cost TLC = THP x LCH
  - Operational Cost OPR = 24% of TLC

TASK 3 - TOTAL COST OF CENSUS AND SAMPLE SURVEY AT NATIONAL LEVEL IN EUROS



Census
Total EU cost
12,7 Million Euro



Survey Sample
Total EU cost
190.682 Euro

TASK 3 - CONCLUSIONS

- The selection of data collection method should be determined by research needs and capabilities
- Data collection systems that rely on voluntary reporting will almost certainly fall short of a complete census
- Most countries currently appear to employ a voluntary approach to data collection
- The fire data collection systems examined in this research appear to be generally regarded as census systems of data collection
- Mixed efforts by national programs to encourage and evaluate compliance
- It is critical that effort be made to identify the extent of missing data and the patterns it takes if data on fire incidents is to be considered reliable
- The impact of missing data is likely to be especially problematic if it fails to account for differences in the population that impact fire experiences
- It is important that the implementation of fire data collection systems include plans for data quality checks and procedures for handling missing data
- The cost of implementing a comprehensive data collection system will be greatest in countries that have the least experience and fewest resources

TASK 3 – FINAL REMARKS

- Data collection systems should be designed with sustainability in mind
- Overly ambitious and detailed data collection systems may tax the patience of participants and undermine data quality.
- Align data collection content with realistic policy goals and use data to promote safety interventions and practices.
- Use data to chart and publicize trends, demonstrate the utility of data collection, and build public recognition and support.