» Data Collection, a shared responsibilities approach
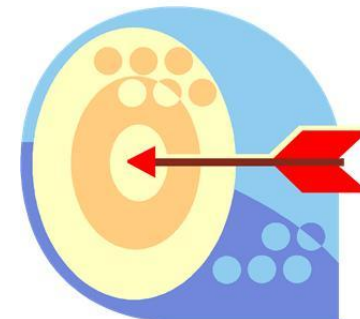
Facilitation of data transfer from enterprises » to NSIs

Luísa Pereira

Luxembourg - 22 May 2013

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

Data collection Objective

Automate data collection process

Web Service XML

Upload XML

Webforms (2005)

Paper

PSI
Information
System

Automation

PAPER
slow, laborious,
error prone

Webreg

WEBFORMS
fast, laborious,
less errors, save costs

webinq √

Statistics Portugal

Survey Management System

SIGINQ

XML
fast, easy,
improve quality, save costs
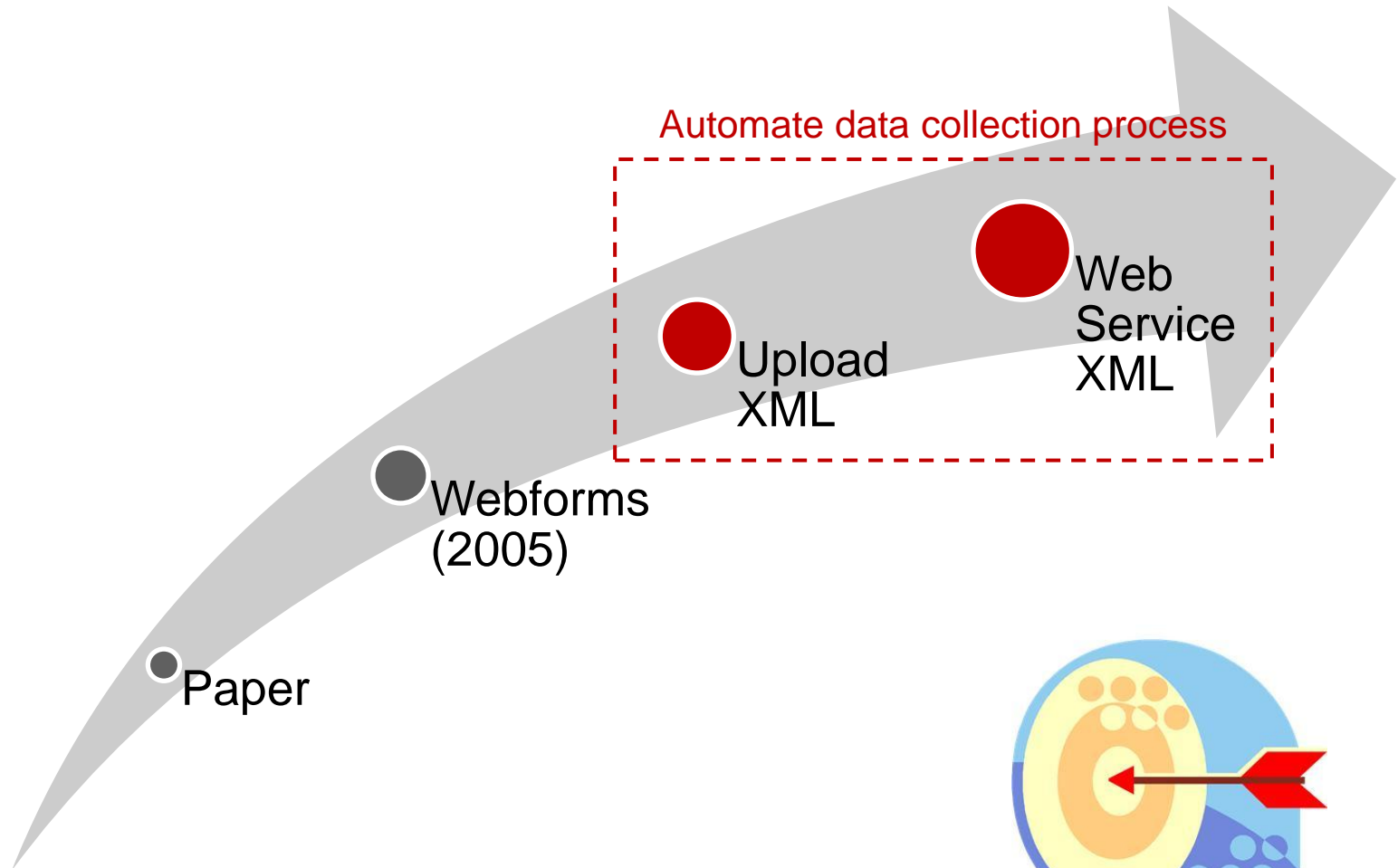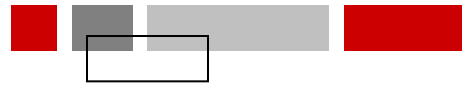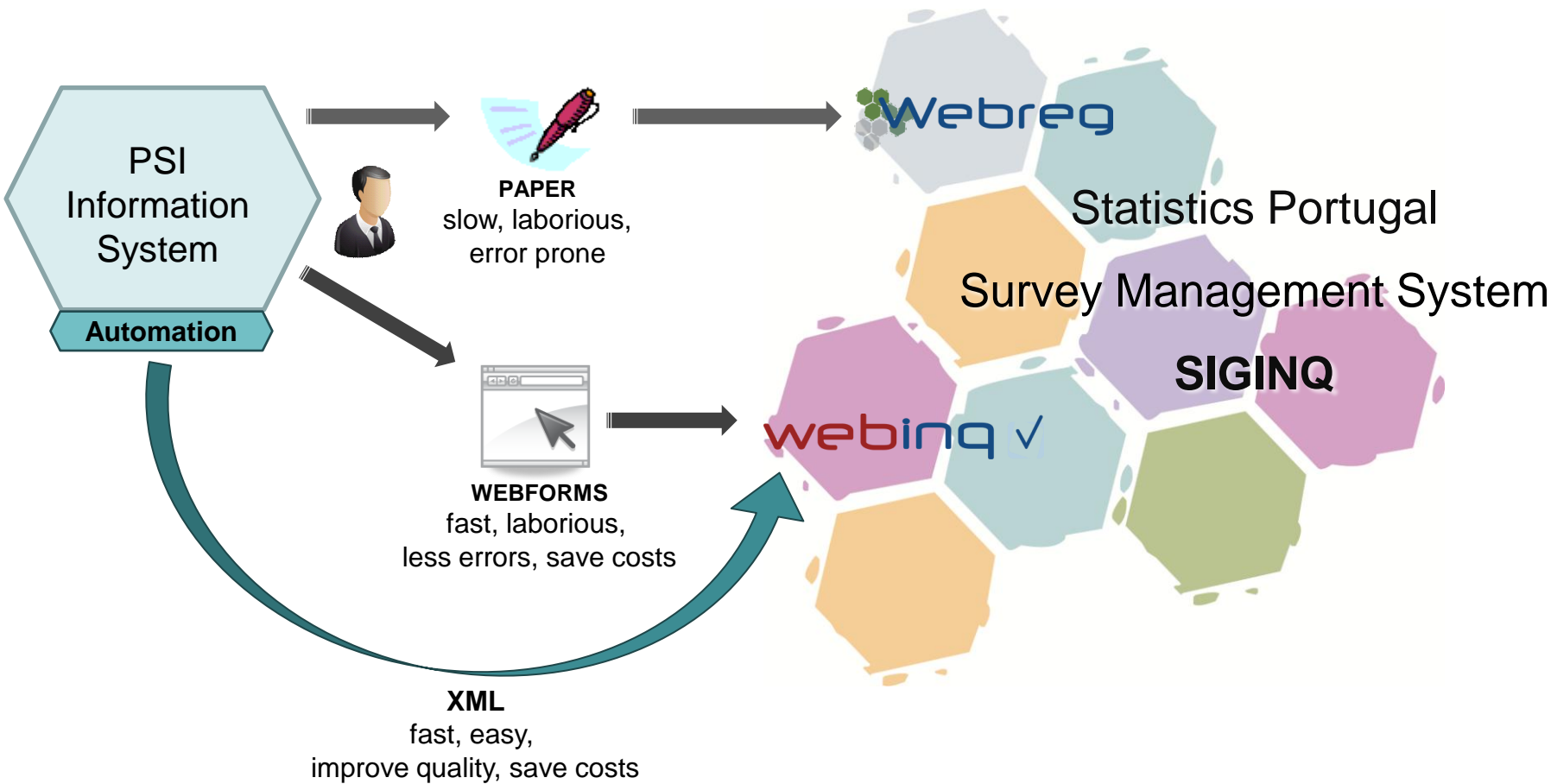
PSI = Provider of Statistical Information

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

Automated data collection
Reasoning and expectations

- **Reduce actual reporting burden**
  - After some initial development work, data submission should be effortless and on time

- **Reduce workload of data collection teams**
  - Data collection teams should equally benefit from automated data collection
    - because answers have guaranteed quality and are incorporated automatically

- **Reduce errors on data input**
  - Errors are expected (and fixed) during the test period
  - Machines are good at repetition, so there should be no errors (typos, misunderstandings, wrong code lists)

- **Reduce overall operational costs**
  - Once set up, the cost of operation is lower for both PSI and receiver

- **Adopt reusable methods, enabling future use in other domains**
  - By following the generic and integrated design of our existing system, we expect the cost of implementing the same method on surveys for different domains to be minimal

- **Minimize impact on current architecture**
  - To avoid the redesign of large portions of an already large and complex system, integration points should be carefully picked

Shared Responsabilities

- **Automation is not for everyone**
  - It does not completely replace paper and electronic questionnaires; they will still be the best way for some PSIs

- **Design with the right audience in mind**
  - Focus design on the use cases of the PSIs that most benefit from automation. In doing so, we expect:
    - Simple and performant solution
    - Short learning curve and low implementation costs for PSIs

- **The key is communication and diplomacy**
  - Getting the cooperation of PSIs is vital to make automation work
  - Ensure PSIs perceive the benefits this approach brings them
  - Show full availability to help clear any doubts
  - If respondents aren't engaged and committed we'll be building an elaborate device that will not be used - therefore useless

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

Choosing the right candidates

- Not recommended for surveys that change frequently:
  - Questionaires
  - Samples
- Recommended for surveys with low periodicity (weekly, monthly, ...)
- Recommended for surveys where PSIs are known to use a 3rd party information solution
  - Software houses will be glad to provide this new feature
  - A single implementation in 3rd party software enables adoption by several PSIs
- Recommended for PSIs that:
  - Already have an information system
  - Are willing to invest in order to reduce the costs of answering surveys
  - Deal with high volume of information
  - Have high reporting burden

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL
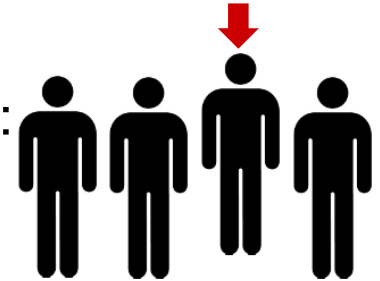
- Too much automation leads PSIs to forget to check the quality of the data

- Answers to open questions, such as justifications may be missing due to their inexistence in the PSIs system

- Since all data is valid, "looks" good and is reviewed by less human eyes, errors can go undetected and are harder to find

- Re-submission of data can lead to waste of hardware resources or - even worse - coherence problems

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

- **Platform independence and easy implementation**
  - are guaranteed for any system PSIs may have, by using plaintext XML

- **Guaranteed future proof**
  - Plaintext XML will always be easy to produce and process

- **Support for unlimited complexity while providing reasonable reading**
  - XML structured nature supports unlimited complexity while providing reasonable reading for machines and humans alike
  - Theoretical support for:
    - Multiple surveys
    - Multiple periods
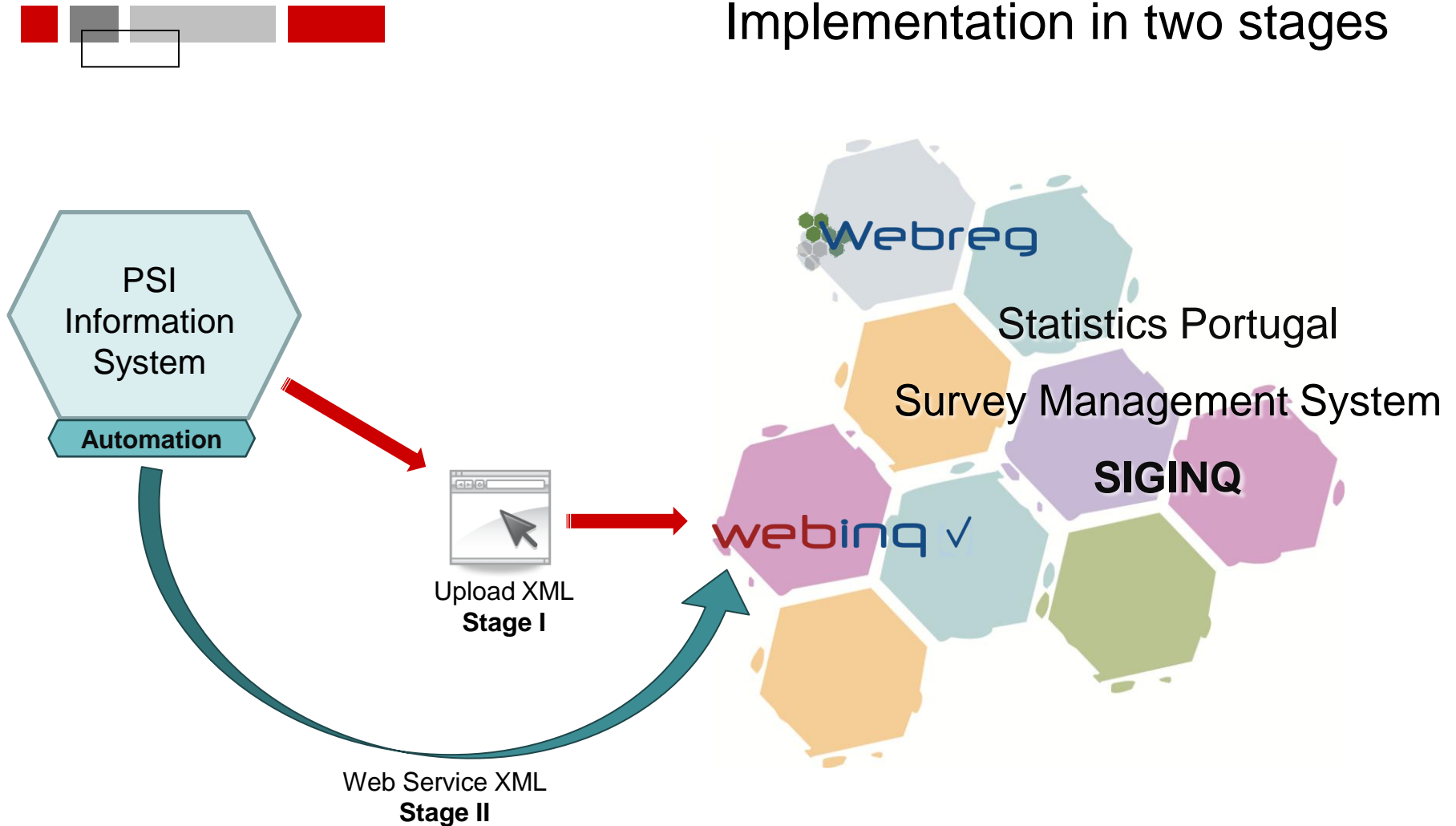   - Multiple statistical units

<XML/>

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

- **Public XSD – early (and easy) basic data validation**
  - Providing a public XML Schema Definition (XSD) guarantees data wellformness
  - Allows for basic data validation before submission from PSIs

- **Simplified XML – easy adoption**
  - Easy, simple XML format reduces the implementation costs for PSIs
  - The chosen schema is similar to our internal data architecture, making implementation and integration easier

<XML/>

Implementation in two stages

PSI Information System

Automation

Upload XML
Stage I

Web Service XML
Stage II

Webreg

Statistics Portugal

Survey Management System

SIGINQ

webinq √

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

- **Initial development**
  - Design and implement the ability to consume (read) XML files with answers
    - Save the extracted data into our data repository
    - Return feedback, either reporting the errors found or displaying a success message

- **Test and tune**
  - To fine-tune this feature, an upload function will be available on a private website accessible only to a few software houses
    - This will allow them to start submitting sample XML answers in order to test the internal integration and assess the difficulties that may be encountered

- **Deploy**
  - When ready, an upload function will be available in a private area of Webinq (eletronic data collection portal) for a specific survey

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

- **Evolution**
  - Allow external systems to interact with our data collection system

- **Web Service**
  - We plan to offer a simple Aplication Programming Interface (API) exposed as a Web Service

- **Simple and secure workflow**
  - Allowing for authentication and authorization, in addition to all the functionality provided by the upload method previously presented
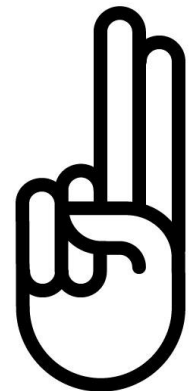
- After some consideration, accommodation-surveys were chosen as the best fit for a real world test:
    - Accommodation-statistics have
        - Monthly frequency
        - Stable questionaire
        - Stable sample
    - Large hotel chains have an information system in place
    - Surveys ask for data already present on the PSI's systems
    - Data collection team showed interest in participating in this test
    - Administration Board agreed that accommodation statistics were a good candidate
    - Portuguese Hospitality Association – AHP, showed interest in being our partner, mediating the interaction between hotels and software houses to implement this automation

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

» Data Collection, a shared responsibilities approach

Facilitation of data transfer from enterprises «
to NSIs

luisa.pereira@ine.pt

Luxembourg - 22 May 2013

Thank You!

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

»15