



Big data: a new world of opportunities for software services

Julie Marguerite

THALES

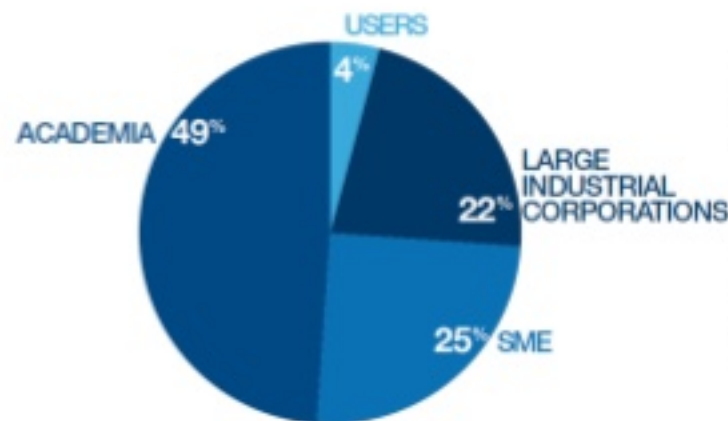
April 9th 2013, Dublin

A mission

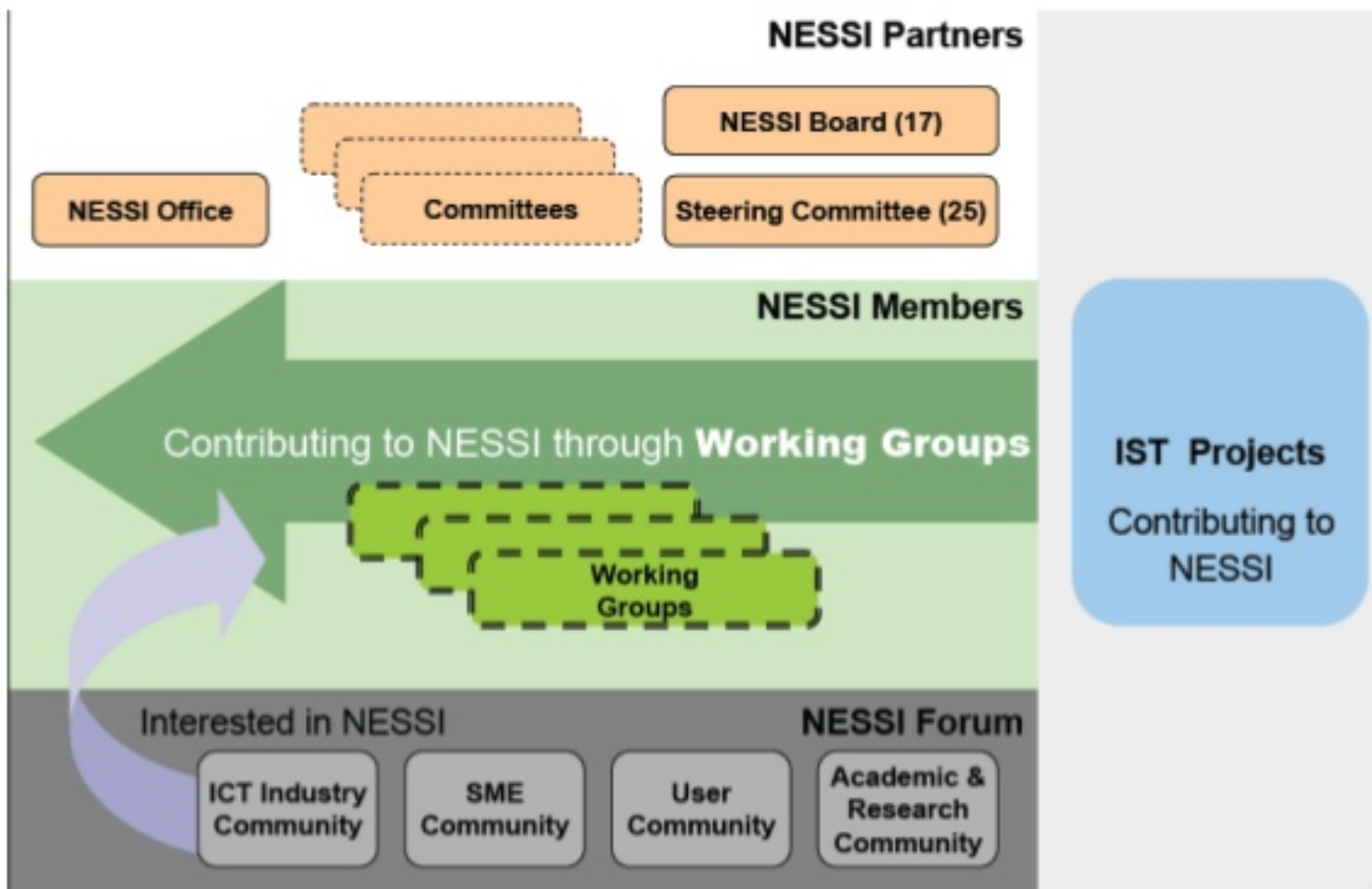
- **Framing the future** of the service oriented economy
- Guarantee **competitiveness** in Europe through ICT innovation
- NESSI partners have recognised that no single company can or should harness the power of such an environment.

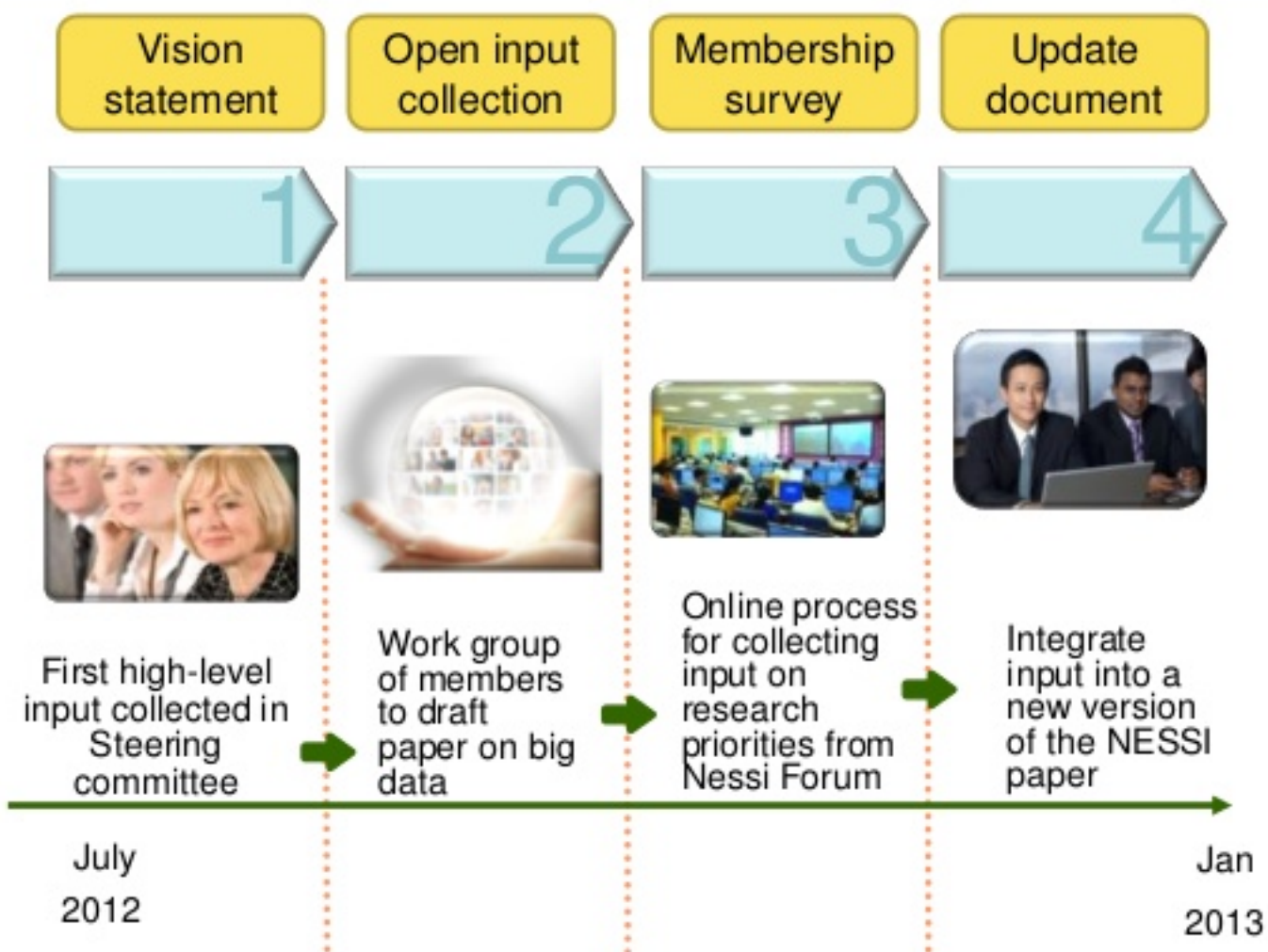
The challenge

- Transform the internet to **combine technologies, services and usages**
- Ensuring that services can be provided to all, citizens and businesses alike, in safe, secure, reliable, extensible, scalable environments.



NESSI represents the software and service IT industry for the members and partners versus Commission and other stakeholders on future technology, industrial and economical/business developments in the domain.





Digitalization led to an explosion of data of all kinds. Generated by:

The Connected World

- Information provided by devices, sensors, objects of any kind increasingly interconnected via the Internet.



Infrastructures

- Intrinsically each transport systems produces considerable flows of data



- New storage architectures and cloud computing: falling storage and processing power prices**
- Awareness of the importance of data value**

People

- Changing lifestyle
- Identity, administrative acts, transportation, banking, social networking...
- Growing internet shopping



Industries

- Increased collections and 'repositories' of data accessible by many industries who could not or did not think possible to store massive information



▶ **In this new context, 2 trends exist...**

- ▶ The real progress that comes with "Big Data" in terms of **processing applications that used to be totally limited** and constrained by issues of "scalability" or storage capability, and become, therefore accessible for analysis.
- ▶ The **extension of some data analysis process** already currently widespread, using proven techniques, but which ultimately benefit from the new capabilities offered by the "Big Data" context, to improve them significantly.

▶ **.... Which lead to 2 big data approaches**

- ▶ Handling large amounts of data is not sufficient to characterize an application or processing of "Big Data", it is also necessary to show that we need an almost exhaustive processing of data and we really need to analyze all these collections together.

Fundamental Big data

- ▶ Complexity of the problem to be solved is simultaneously at the size of data to handle and at the combinatorial calculations to perform
- ▶ Requires: algorithms operating or storage that take advantage of the structure of the data being analyzed.
- ▶ Example: identification of communities for cyber, e-reputation...

Big data by extension

- ▶ Manage business applications without having to make full use of all available data, even if they are in large number.
- ▶ Requires: big samples and improvement of existing process analysis.
- ▶ Example: reactive marketing offers for mobile phone customers

- ▶ Analysis of data (big data analytics) allows to extract non-trivial information which would not be accessible otherwise
- ▶ Some scalable techniques to investigate in a big data context:
 - ▶ Association rules
 - ▶ Genetic algorithms
 - ▶ Cluster analysis
 - ▶ Learning methods
- ▶ Pure analysis of huge repositories of data without a tight coupling with decision support approaches and tools is of minor interest.

- ▶ **Production of data (capture, input, etc.)**
 - ▶ Contextual awareness to decrease complexity
 - ▶ Content validation
- ▶ **Storage**
 - ▶ Distributed storage of data sets
- ▶ **Treatment / analysis / visualization**
 - ▶ Developing scalable algorithms for processing imperfect data in distributed data stores;
 - ▶ Modularization of large graphs
 - ▶ Streaming processing i.e processing of data streams on the fly often in real time.
 - ▶ Developing computational techniques and software tools for analyzing large volumes of data, both semi-structured (e.g., tabular, relational, categorical, meta-data) and unstructured (e.g., text documents, message traffic).
 - ▶ Creating effective human-computer interaction tools for facilitating rapidly customizable visual reasoning for diverse missions.

- Growing importance of "data scientists" and multidisciplinary team in the business.
- Behind is the issue related to the **implementation of new curricula**.
- New education initiatives **related to "data science"** such as: Coursera, Udacity, EdX, etc.. quickly democratize skills that were previously difficult to promote outside of academia.
- NESSI encourages the set up of education programs on data science. The education challenge is not only to teach students fundamentals skills such as statistics and machine learning, but also appropriate programming ability and visual design skills.

- **Mutualisation of Data needs trust and confidence to facilitate the sharing with others**
 - At any level of a process involving Data treatments or Data exploitations security, privacy ,trust must be included end to end during the algorithmic process of Data transformation.
 - Guarantee to all users at all time a protection of their data (integrity and usage) while allowing the use of these data to improve the quality of the services provided.
 - Manipulating information dealing with personal data needs a controlled process.

- **New approach to trust and security**
 - Data flows everywhere and autenticate, abnormal patterns are traced and isolated
 - Strategy is no longer to close access but to restrain damages.
 - Trust label associated with the data and its usage
 - Learning techniques for business oriented threats mitigation

- ▶ Big data is a game changer in all industries enabling new business scenarios and new way of providing services

- ▶ Paper: Strength in numbers: How does data-driven decision making affect firm performance?
 - ▶ Out of the 179 enterprises studied, those which adopt a decision mode according to data processing, obtain a **5 to 6% productivity gain** that cannot be resulting from any other factors.
 - ▶ MIT , Erik Brynjolfsson, Lorin M. Hitt, and Heekyung Hellen Kim.

- ▶ Managing data in an effective way opens a wide field contributing to the improvement of services such as:
 - ▶ on demand" and context-sensitive transportation strategies,
 - ▶ optimized management of energy demand,
 - ▶ more "holistic" and "preventive" health care approaches,
 - ▶ development of new services such as e-voting, etc.

- ▶ **Cyber security and fraud detection**
 - ▶ Intrusion detection systems
 - ▶ Fraud and abuse management systems
 - ▶ IT network vulnerability analysis
 - ▶ Cyber crime investigation
- ▶ **Spatial and aeronautics**
 - ▶ Classification of stellar objects
 - ▶ Stars cartography in astronomy
- ▶ **Video surveillance and image archiving**
 - ▶ Videos and images search and retrieving
 - ▶ Research of similar scenes
- ▶ **E-government**
 - ▶ E-voting
 - ▶ Taxes management
- ▶ **Smart cities and transportation**
 - ▶ Smart water, traffic, energy management
 - ▶ Automatic billing and dynamic pricing
 - ▶ Ticketing collections analytic exploitations





More on NESSI: www.nessi-europe.com

Download NESSI white paper on big data at:

http://www.nessi-europe.com/Files/Private/NESSI_WhitePaper_BigData.pdf