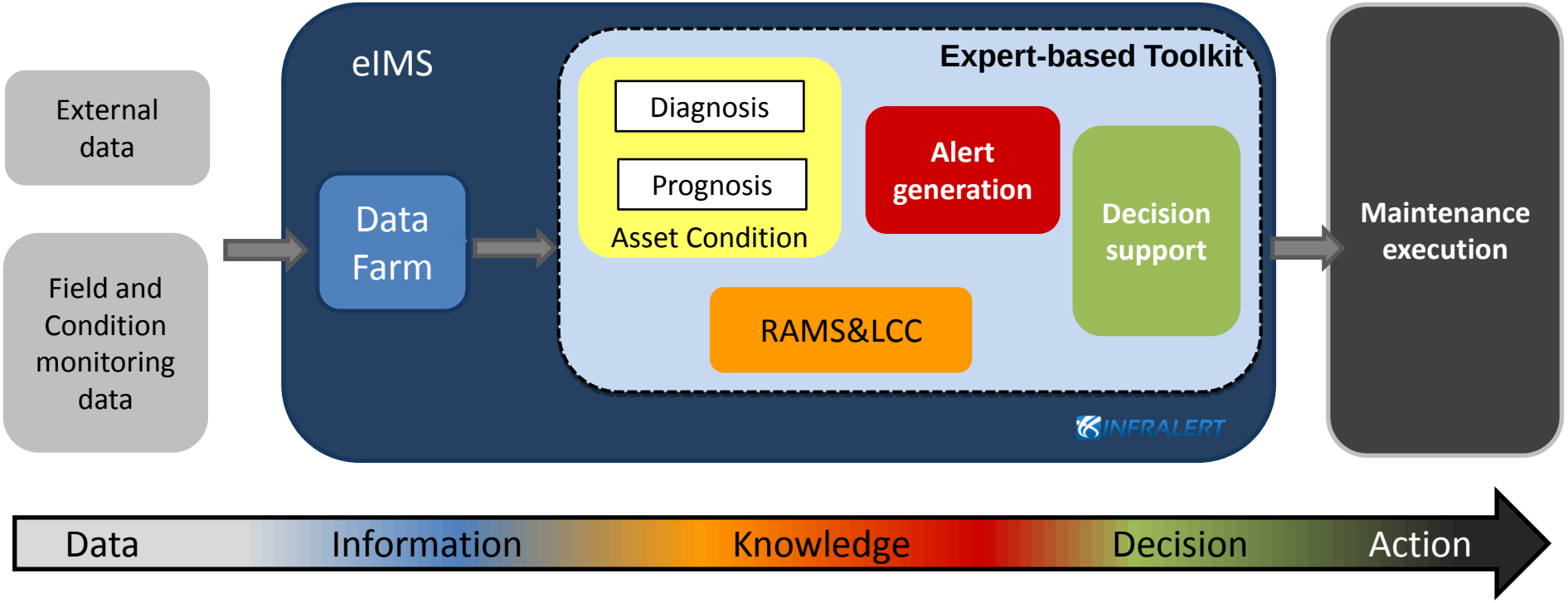
THE DATA FARM

1st Open Workshop, Brussels 16.11.16

Cesare Santanera

GENERAL OVERVIEW

INFRALERT develops an **expert-based Infrastructure Management System** which **coordinates and integrates** all the processes for **Maintenance & Renewal** and support long term strategic investment decisions.



OBJECTIVES

- The prime objective of WP2 – Data Management, is the creation of a Data Farm, innovative from certain points of view, able to satisfy the project's needs.
- A *Data Farm* is basically a data repository. It is an organised data container where stored information should be easily retrievable.

FEATURES

- The principal features that the INFRALERT Data Farm must have are the following:

✓ Ability to store a huge amount of data, even big-dimension data, without compromising processes speed.

✓ Ability to evolve easily over time, i.e. able to incorporate new data types when and if it will be necessary.

FEATURES

- ✓ Ability to be easlily accessible, in open and documented ways.

- ✓ The Data Farm to be developed, as opposed to a plain database, is an engine able to collect, transform and assess big data, while it still is scalable and manageable. These conflicting needs are addressed by taking advantage of the linear nature of the infrastructure to describe.

RELATIONAL DATABASE

- The core of the Data Farm will be a relational database. This choice over a non-relational db offers a series of advantages.
- SQL is an excellent language to search for data: the searches are based on relations construed on the data having a meaning for the users.

DATA FARM FEATURES

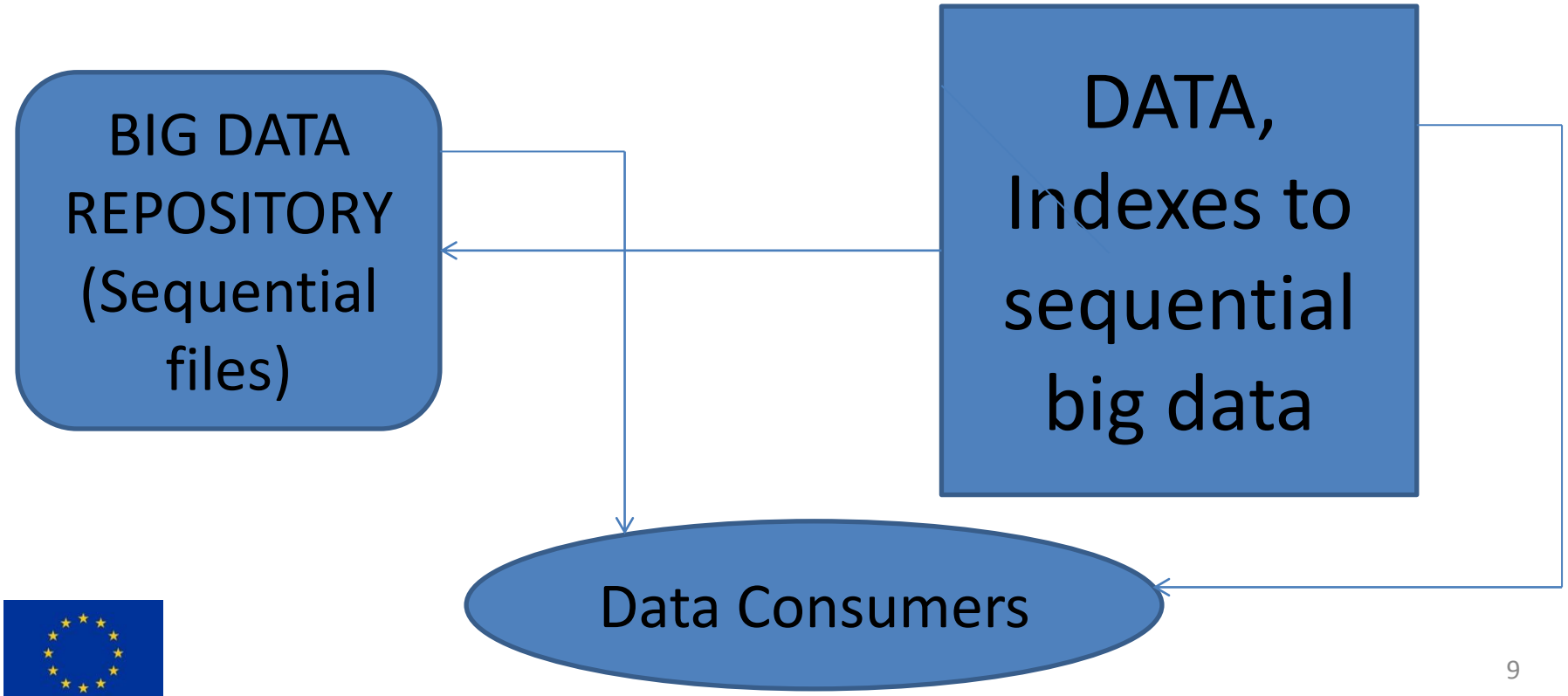
- SQL advantages are delivered at the expense of access speed and not optimised storage size. In fact, a relational database is very inefficient for big volume data. (e.g. video inspections and cross sections).
- However, the obvious topological model of "linear" infrastructure is an edges and nodes structure: a relational database is very well suited to represent this structure.
- The bulky data are linked to the edges, not to the nodes, and these data are intrinsically sequential.

DATA FARM FEATURES

- A mixed structure, featuring the advantages of both the relational and the sequential worlds, is therefore necessary.
- It means that the INFRALERT Data Farm cannot be only a ‘simple’ relational database.

DATA FARM STRUCTURE

- Data storage will be split in two containers: the relational database and the Big Data Repository, that will store big dimension, sequential data.



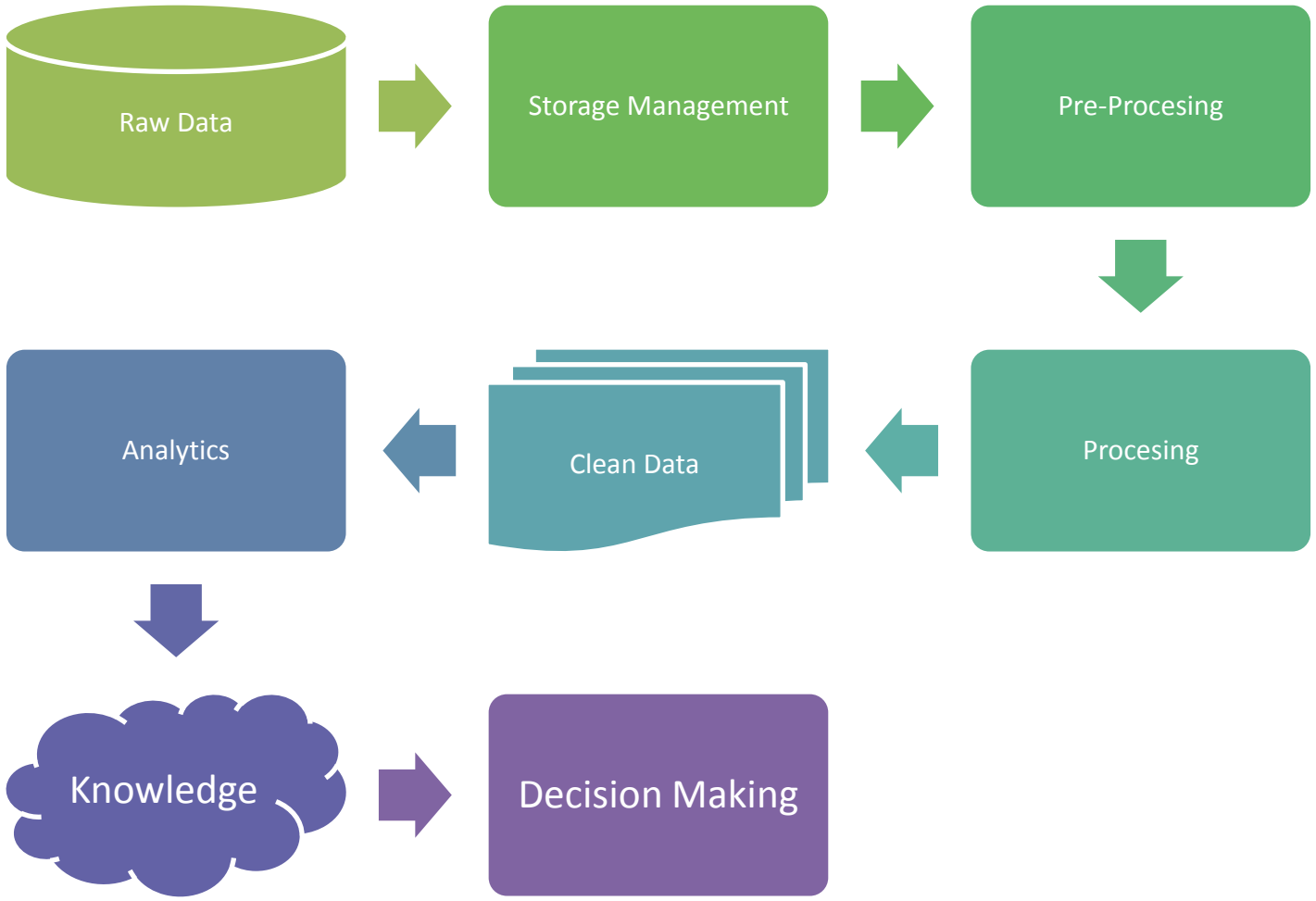
STRENGTH POINTS

- It is important to notice that currently existing Data Farms are built for asset management, rather than for measurements from the field, hence are inefficient in handling big data such as videos. The Data Farm to be built within the INFRALERT project will be able to store high-volume data and will also be holistic to the extent of including all types of data in a consistent architecture.

DATA MANAGEMENT

- Data management techniques are the ones involved in the process of transforming (big) data from original format (raw data) to computer formats and it progresses with applying (big) data operations towards achieving decision-making. The goal is usually to obtain a clean data repository to which data analytic techniques are applied in order to extract knowledge for decision making. This process is schematized in the following figure.

DATA MANAGEMENT



INNOVATIONS

- The INFRALERT Data Farm, in contrast to conventional databases, will store big volume data and will still be open, portable and scalable.
- SQL is an abstraction layer making the applications portable: everything will be done in order to make it easily installable on a number of popular databases, from MySQL to Oracle.



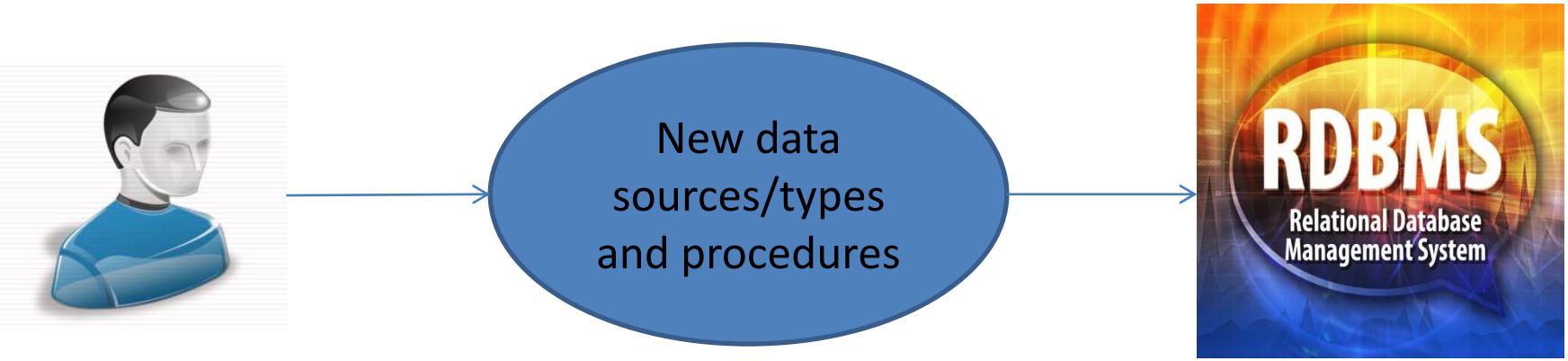


KEY FEATURES

- User expandable.
- Direct access to the data possible for the competent user.
- Compliant with the standards.
- Data access speed.

USER EXPANDABLE

- The users shall be able to add data sources and data types. Every user/organization shall be able to use a manual to update the system.
- Once new data are added, they must be used. Thus a programmer shall be able to add procedures using the newly added and the previously existing data: it will not be necessary to call the system manufacturer for every change/upgrade.



DIRECT ACCESS

- Direct access to data is possible for the competent user.
- The users can interrogate the data base by building their own queries.
- The architecture allows the competent user many ways to add his own algorithms. The main ones are: a plug-in on the server, a brand new client, seeking the data on the server via any of the available methods.

MY PROCEDURES

COMPLIANT WITH THE STANDARD

- Some formal or de facto standards exist (especially for railways) .
- It is important to incorporate these standards and expand them consistently where the standards are not enough to model the infrastructure to the desired detail level.

DATA ACCESS SPEED

A very important point allowing us to think our ambitious goals are atteinable is also:

- The very nature of the data coming from the measurements on linear infrastructures. These data are intrinsically ordered sequentially, along the space coordinate.
- This "a priori" knowledge can be incorporated into the system architecture, to create an access engine very efficient both in terms of speed of access and in terms of storage volume.

INFRALERT

www.infralert.eu

Forename NAME

Company
E-mail