Provenance for the safety control of large civil engineering structures

A Dam Safety Control Scenario

José Eduardo Barateiro
jbarateiro@lnec.pt

20th April, 2009 – Use Cases for Provenance Workshop

Outline

- Motivation
  - Digital preservation
  - Data quality
  - Interoperability
  - e-Science
- Dam safety control
  - Concepts
  - Archive example
  - Information Lifecycle
  - Main challenges
- Current information system
- Conclusions

Motivation (1/3)

- Digital preservation (DP)
  - Digital contents are unstable, requiring continuous management to remain interpretable.
  - Each type of digital object has its own specific requirements.
  - Several communities like biology, engineering or physics, manage large amounts of valuable and heterogeneous scientific information (e.g., data sets, mathematical simulations and highly specialized documents reporting research work) that must be preserved.
- Data quality
  - Data quality processes can be seen as sequences of data transformations (often modeled as a graph of data operations) that must be applied in order to get data of good quality

Motivation (2/3)
Digital Preservation and Interoperability

- **Digital preservation** aims at ensuring that digital objects remain retrievable, accessible (technical metadata) and usable over a long period of time. Authenticity and integrity must be guaranteed.
- **Interoperability** ensures that users, systems and processes can exchange and share information.
  - Technical, Semantic, Human, Legal, International

Digital preservation vs. Interoperability?

- DP -> **time** dimension (communicate with the future)
- Interoperability -> **spatial** dimension (communicate with others)

Motivation (3/3)

Provenance *in e-Science* scenarios

- The emerging collaborative environment of the scientific community, with its associated services and infrastructure, assume data sharing, interoperability and, in some scenarios, impose strong requirements for long-term preservation and data quality.

- e-Science, involving intensive computation and massive data sets, is growing strong.

- **Provenance information is crucial for preservation, data quality and interoperability purposes**

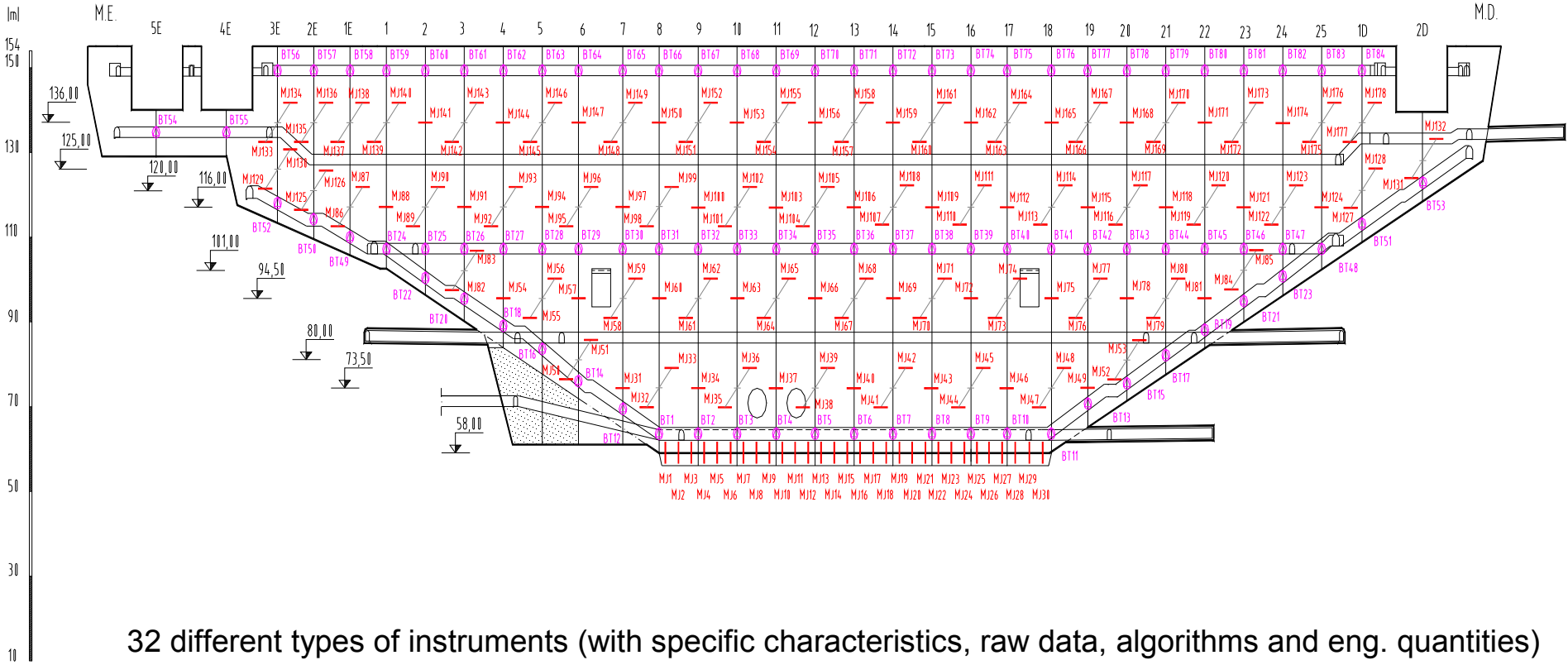
Dam Safety Control (1/8)

Concepts

- According to the Portuguese Dam Safety Legislation, LNEC is responsible to keep an electronic archive of data concerning the dam safety.
- Dams monitored by instruments (e.g., plumb line).
- Raw data (*readings*).
  - Manually collected.
  - Automatically collected.
- Raw data transformed into engineering quantities, by specific algorithms.
- Geodetic observation.
- Mathematical simulations.
- Physical models.
- Visual inspections.
- etc.

Dam Safety Control Scenario (2/8)

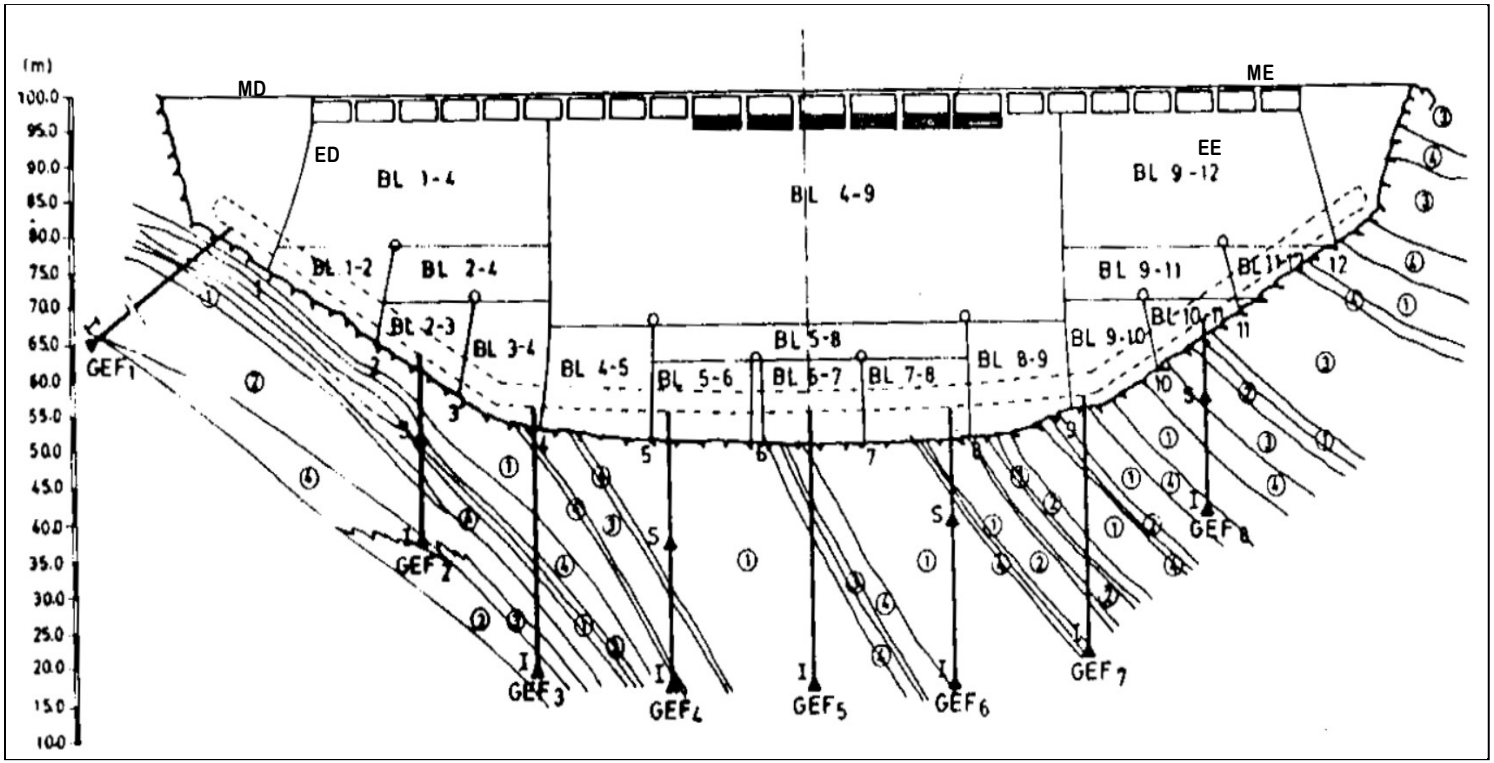
Concepts - Instruments



32 different types of instruments (with specific characteristics, raw data, algorithms and eng. quantities)
Hundreds to thousands of instruments per dam (automatic monitoring and manual gathering)
~100 dams monitored

Dam Safety Control Scenario (3/8)

Concepts – legacy data

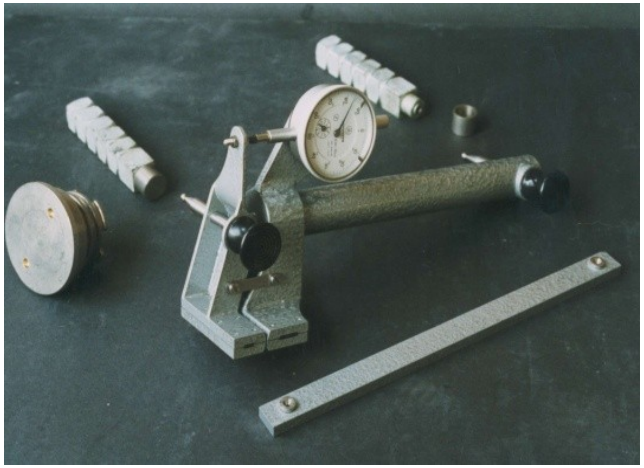


Highly heterogeneous data: e.g., archive must comprise legacy information as project drawings and handwritten observations!

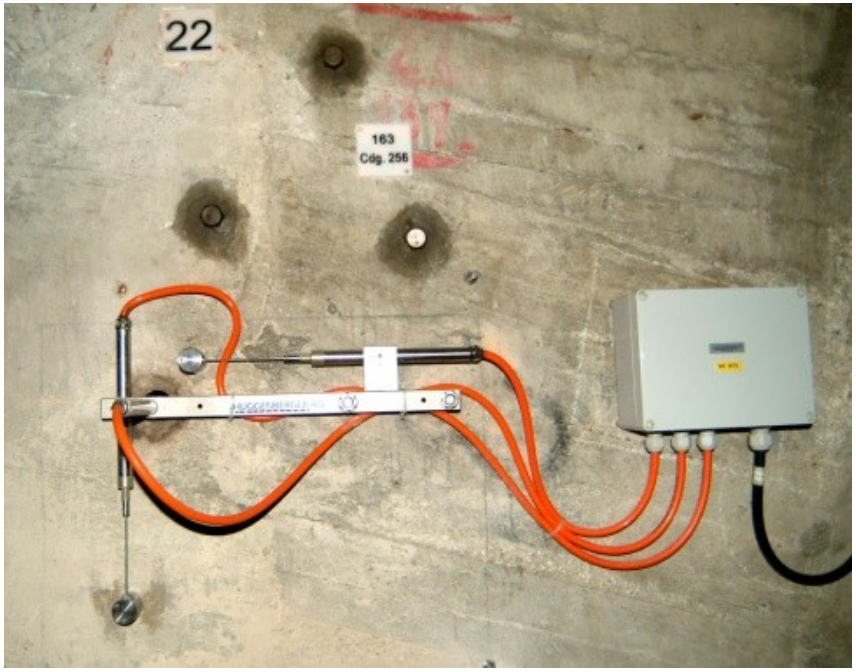
Dam Safety Control Scenario (4/8)

Concepts - Manual vs. Automatic Observation

Manual observation using specific measuring instruments



Automatic monitoring

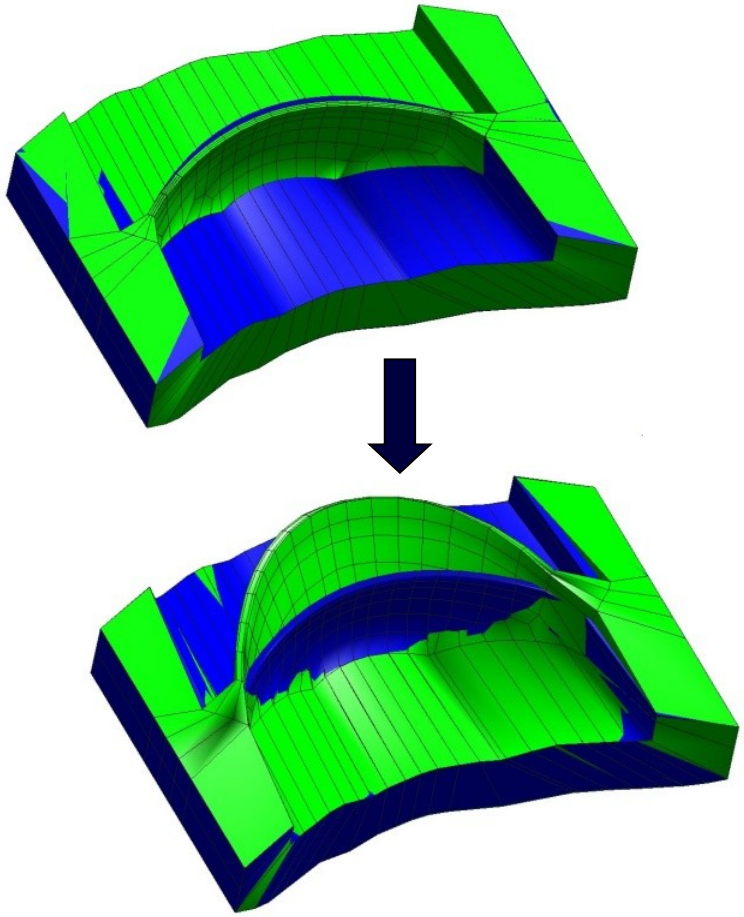


Dam Safety Control Scenario (5/8)

Concepts - Physical Models and Visual Inspections

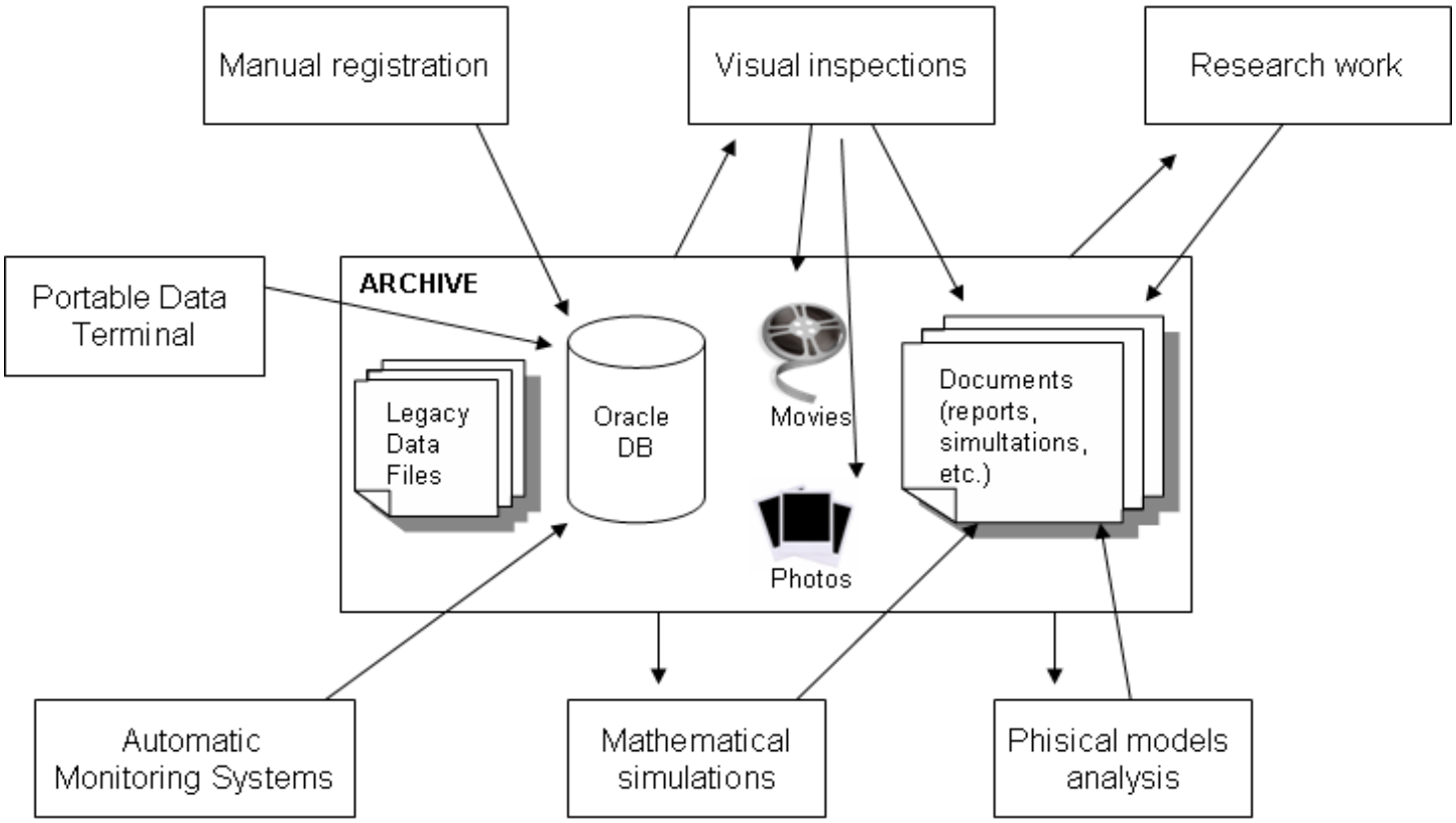
Scaled Physical Models

Mathematical models



Dam Safety Control Scenario (6/8)
Archive example

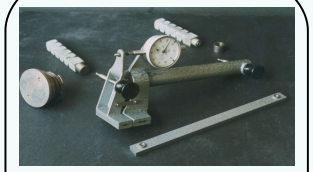
- Archive example



Dam Safety Control Scenario (7/8)

Information Lifecycle example

INGEST

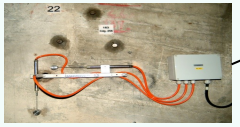


Manual

PDT

Manual

By File



Automatic

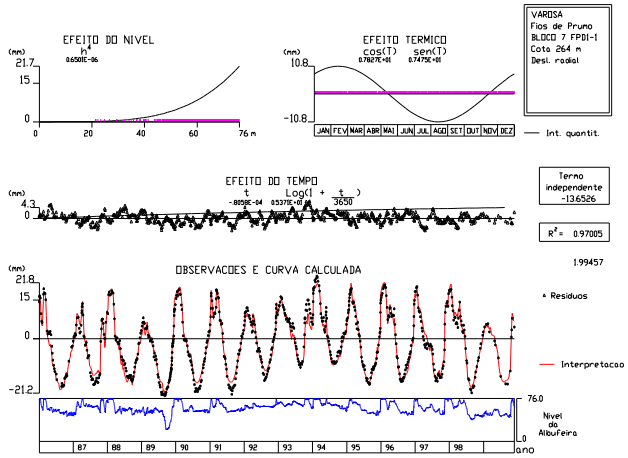
Raw data transformed by proprietary systems

Dam Safety information system

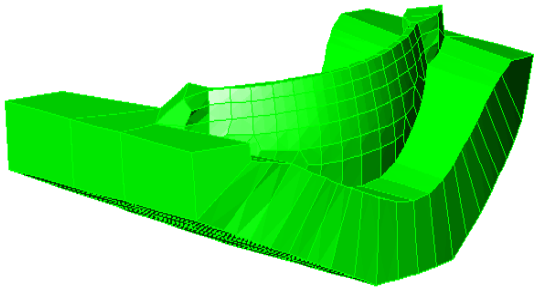
Services

DB

Conversion to engineering quantities



Statistical models



Mathematical simultations, etc.

Dam Safety Control Scenario (8/8)

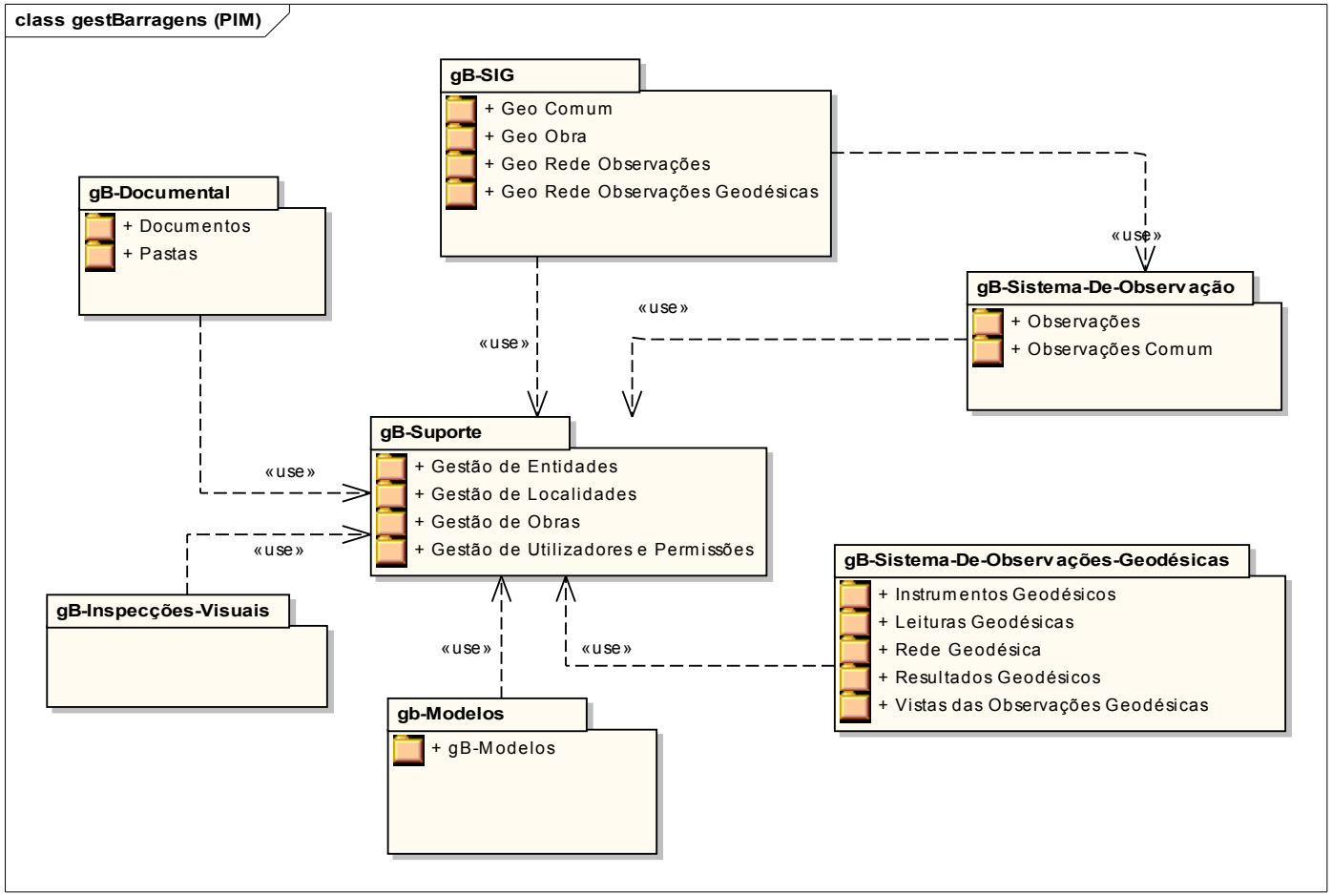
Main issues

- Heterogeneous data
  - Raw data (manually gathered)
  - Raw data (proprietary automatic monitoring systems -> include algorithms, calibration constants, etc.)
  - Calculated data and respective algorithms (depend on current calibration constants
  - Specialized documents in several formats: CAD, MS Office…
- Dynamic datasets (not only appended)
  - dependence on the information system
- Context preservation
- Federative environment (several dam owners capture raw data)

Current Information System (1/3)

gestBarragens – Component Model (in Portuguese ☹)

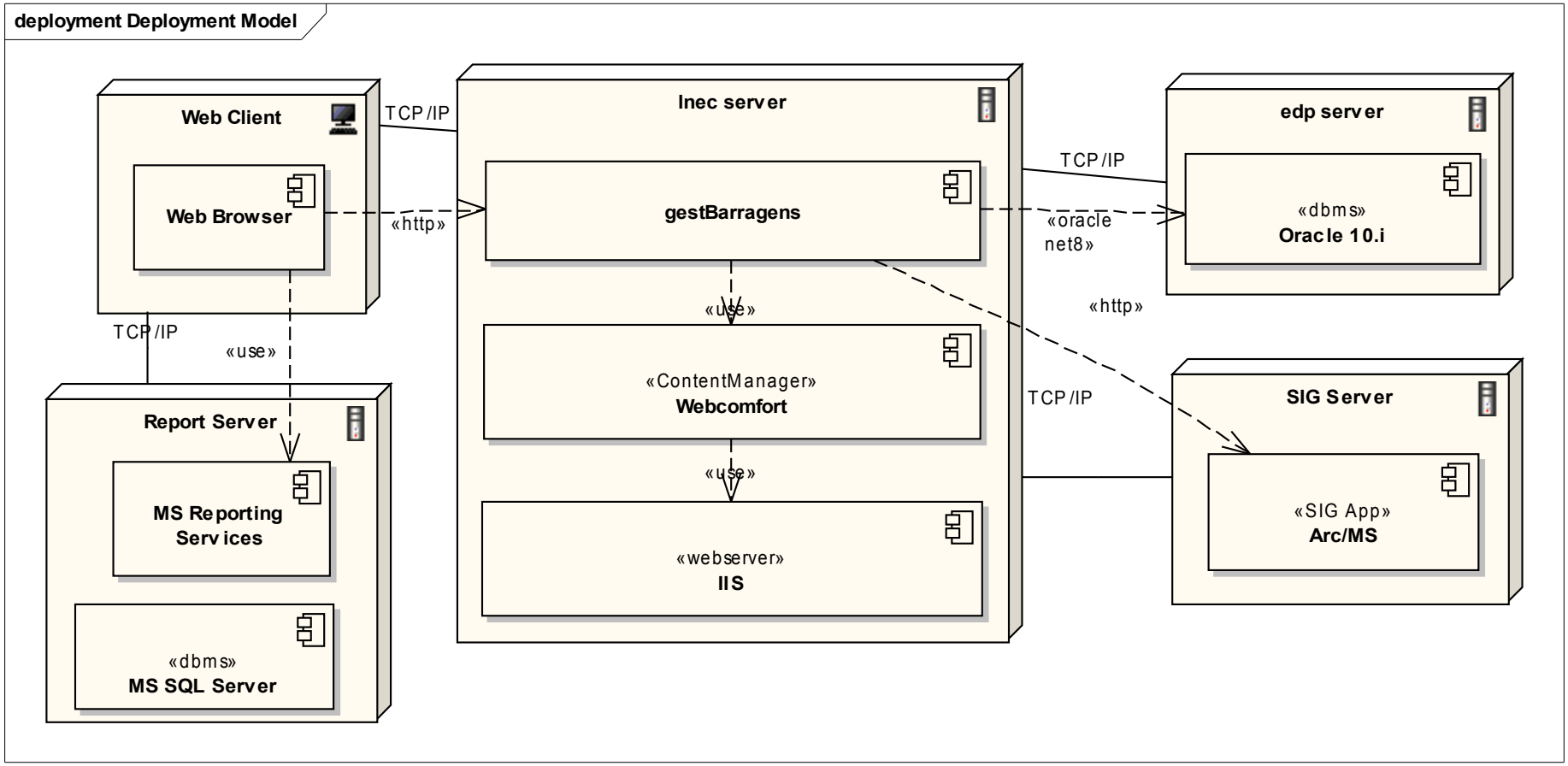
- Modular Information System



Current Information System (2/3)

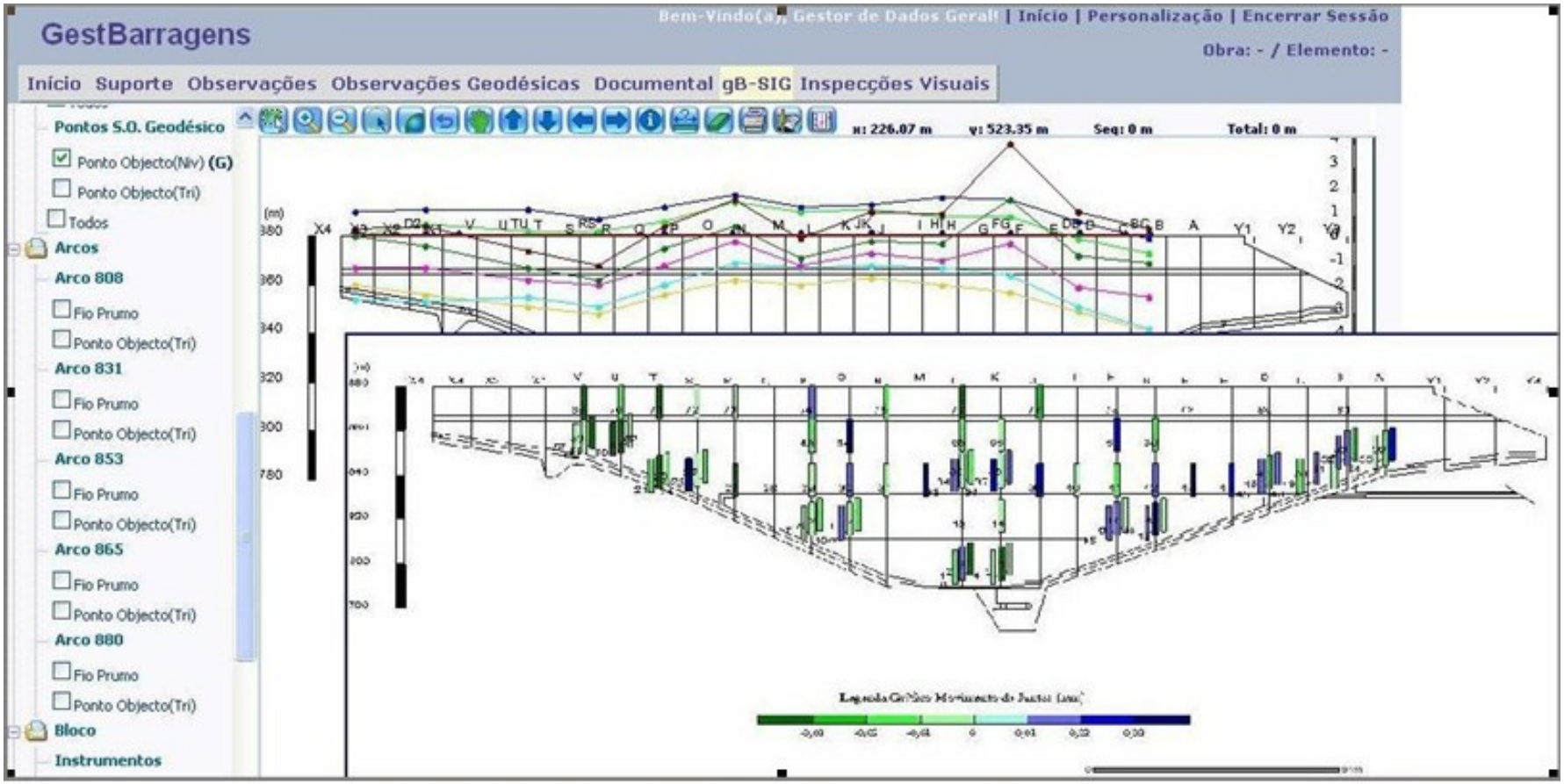
gestBarragens – Deployment Model

- Currently composed by two instances: LNEC and EDP - the Portuguese Electricity Company



Current Information System (3/3)

gestBarragens – Exploring Data



Conclusions

- Civil engineering structures are monitored through the analysis of data captured (automatic or manually) by instruments installed in strategic points of the structures.
- The life-cycle of dam safety data is affected by different processes and systems, often involving different data schemas and file formats
- LNEC and EDP use the *gestBarragens* information system to manage the data collected from dams
- gestBarragens does not handle crucial provenance information required for data quality, preservation and interoperability purposes.
- gestBarragens only controls a specific part of the overall process (e.g., it does not control data capture, simulation models, etc.)