



















### Approach

### • Develop a framework for:

- Providing a formal base for quality evaluation
- Analyzing quality factors and metrics
- Identifying DIS properties that influence quality factors
- Developing quality evaluation algorithms

#### • For each quality factor:

- Analyze definitions, metrics, properties
- Model DIS properties
- Evaluate data quality
- Analyze critical paths
- Identify improvement actions
- Generate probabilistic models for data quality behavior

InCo – Urugu

STIC-AMSUD meeting - 22/7/2008







































# **Examples of instantiation**

#### • Goal 1:

- Improve the quality of students location data (phone number, address, etc.)

	Questions	IS objects	Quality factors
1	Are students' addresses the correct ones?	Student's address	Semantic correct.
2	Are the students' addresses correctly written?	Student's address	Syntactic correct.
3	Are the students' telephones valid ones?	Student's phone	Syntactic correct.
4	Do we have precise students' addresses?	Student's address	Precision
5	Are students' addresses up to date?	Student's address	Currency
6	Do we have all students' addresses?	Student's address	Coverage

STIC-AMSUD meeting - 22/7/2008

InCo – Uruguay

# **Examples of instantiation**

#### Syntactic correctness Address syntactic correctness

#### • Question:

STIC-A

- Are the students' addresses correctly written?

Metrics	Methods
Syntactic correctness Boolean	CheckRule
Address format completion	- student's address attributes
	- address standard format
Syntactic correctness Boolean	CheckDictionary
Address syntax existence	- student's street attribute
	- street dictionary
Syntactic correctness	ComputeDistance
deviation	- student's street attribute
Address syntax deviation	- street dictionary
	- string-edit-distance





llci	Using Obox-Eoundation						
Defining quoti-roundation							
Goals : Goals catalog :: Quality Model : : Quality catalog :: IS Objects :: : : Objects catalog :: : : : : : : : : : : : : : : : : :							
Goal edition							
Name Improve clients data quality							
Des	scription	Improve clients data quality in order to reduce mailing problems					
Dat	te						
Que	estions	Description         Scenario           how good is address related         Factor         Semantic Correctness         Change factor           Metric         Semantic correctness r;         Delete metric         Metric         Metric         Delete metric           Metric         Column         Clientdb people address         Choose         Add metric         Metric         Metric					
_		Add auestion Delete auestion Save Execute					
STIC-AMS	SUD me	eeting – 22/7/2008	InCo – Uruguay				













#### Past applications:

- Facultad de Ingeniería, Uruguay
  - Domain: Student's data and their activities
- ACCSA (consultant cabinet)
  - Domain: Supervision of services to customers
- Crédit Agricole Uruguay
  - Domain: Customers data and accounts
- Institut Pasteur de Montevideo
  Domain: Biological experiments (microarray data)

#### Current projects:

- Microsoft Research project
- Domain: Genetics
- CSIC project
  - Context: Web warehousing applications (web-services)

InCo – Urugua

STIC-AMSUD meeting - 22/7/2008













#### **Motivations**

- **Research problems** 
  - Quality management during query evaluation
  - Quality management during DIS design
  - Quality management reacting to quality changes

#### Our contributions and challenges ۲

- A framework for quality management
- A mechanism for defining quality measurement methods
- **Applications**
- Conclusions

STIC-AMSUD meeting - 22/7/2008



InCo – Uruguay

InCo – Urugua



- Implementing further functionalities
- Incorporating quality of services
- Further applications are planed:
  - Validating the instantiation of previously defined methods
  - Enriching libraries
  - Collecting quality values from real application datasets

#### • This development constitutes a first step for exploring:

- Interdependencies among quality dimensions
- Quality patterns
- Quality improvement techniques

STIC-AMSUD meeting - 22/7/2008

InCo – Uruguay