
DATA VALUE ENHANCEMENT FOR AMAZONIA DRAINAGE DATASETS



MONICA DE MARTINO (CNR-IMATI)

CONSIGLIO NAZIONALE DELLE RICERCHE

ISTITUTO DI MATEMATICA APPLICATE E TECNOLOGIE INFORMATICHE



Sergio Rosim (INPE)

Instituto Nacional de Pesquisas Espaciais

São José dos Campos, Brazil

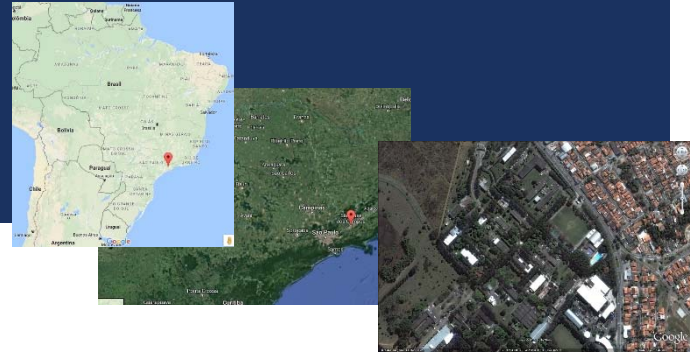
OVERVIEW

- **Background**
 - Water and Drainage datasets
 - TerraHidro (INPE)
- **Objective**
 - Drainage Data of South America transparency
- **Strategy**
 - Overall Strategy
- **Preliminary activity**
 - Data creation with TerraHidro
 - Workflow description of data provenience
 - Open data publication
- **Conclusion and next steps**

BACKGROUND

- World fresh water is a fundamental limited resource.
 - South America region preserve the quantity and quality of significant world's water amount (main distribution:28.3% of water and Brazil holds the 12%)
- Environment water based calamities (flooding, atrophic action)
 - Needs of prompt responses from the governmental bodies both to prevent (forecast) and/or recover (knowledge) environment calamities.
- Need of provision of accurate and update territorial information
 - The **drainage dataset** (e.g. drainage network extracted from raster Digital Elevation Models (DEM), watershed or basin) quick available on the Web

BACKGROUND



- **INPE and TerraHidro**

- The National Institute for Space Research (INPE) in Brazil has the mission of producing and providing drainage datasets to support decision activities to cope with environmental issues.
- **TerraHidro** a distributed hydrological model system aimed at extracting drainage networks and watershed datasets of every earth region starting from a raster DEM. (<http://www.inpe.br>)

OVERALL OBJECTIVE

- To enforce the publication and sharing of the drainage dataset to enable researchers, governmental institutions and private organizations to benefit from their usage in their applications
- To transform of the actual *on-demand* approach in an overall transparent system capable of supply detailed information of South America territory, semantically integrated with other territorial description

ACTIVITY

- To identify a methodology aimed at
 - INPE's **drainage dataset sharing**
 - Provision of the relevant information about the different stages of the data processing pipelines for their **reproducibility**
- In a medium-long term, to establish a consolidated **point of access of interlinked geographic data as well as the metadata-related information**

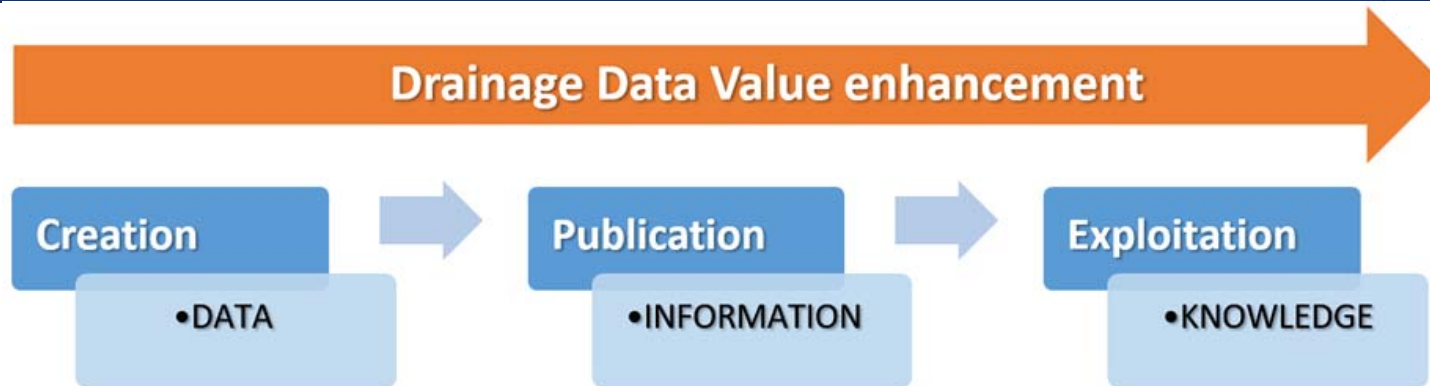
STRATEGY

- **Open Government Data (OGD) principles considering**
 - The implementation of the Brazilian e-government initiative about **data transparencies** according to Law no. 12.527,
 - The strategy of Brazil and the European Commission about “**free and open public access of information**”
 - INSPIRE experience to meet the scientific and technological policy between Brazil and European Commission

APPROACH

- Data Documentation with proper metadata for third party application
 - covering issues such as provenance, quality, versioning (e.g., GEO-DCAT, W3C-DQV, W3C-PROV)
- Data publication on the Web according to the (Linked) Open Data initiative
 - adopting W3C recommendation (Best Practices for (Spatial) Data on the Web)
- Data semantic enrichment
 - considering Linked Data and Semantic Web tools for the semantic enrichment of the Drainage Data

STRATEGY SCHEMA



- **Data Creation:** data processing computations, according to a well-grounded drainage workflow, that extracts several drainage networks and basin datasets from the DEM
- **Information Publication:** data and related metadata on the Web as Open Data
- **Knowledge Exploitation:** semantic integration of drainage data with other datasets through Semantic Web technology, to be exploited at cross-disciplinary level

PRELIMINARY OUTCOME

- **Data creation:**
 - Creation and Identification of data provided by TerraHidro
 - Provision of a **workflow description** to explicit the drainage dataset creation process
 - to identify provenance information, exploitable for reproducibly tasks by third parties
- **Publication**
 - South America and Amazon drainage datasets on the Web according to the Open Government Data (OGD) principles

THE STUDY AREA

- South America region and Amazon basin
 - Raster digital representation by SRTM DEM (30 m resolution)



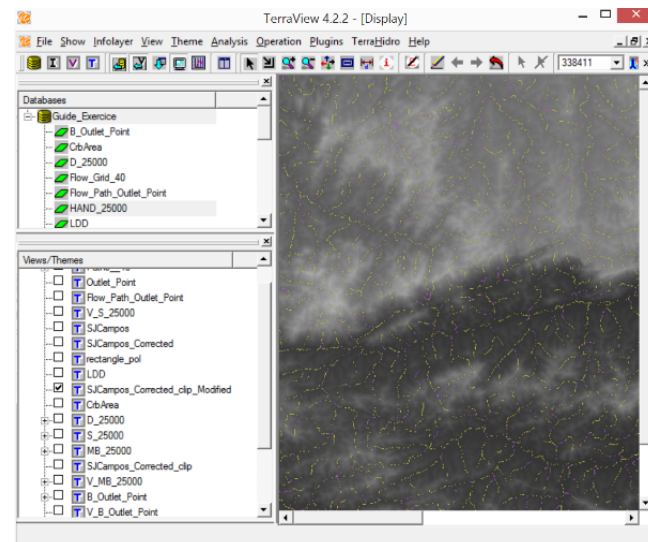
TERRAHIDRO DATA CREATION

WHAT

TerraHidro is a Distributed Hydrological System for hydrographic basin water flow GIS applications.

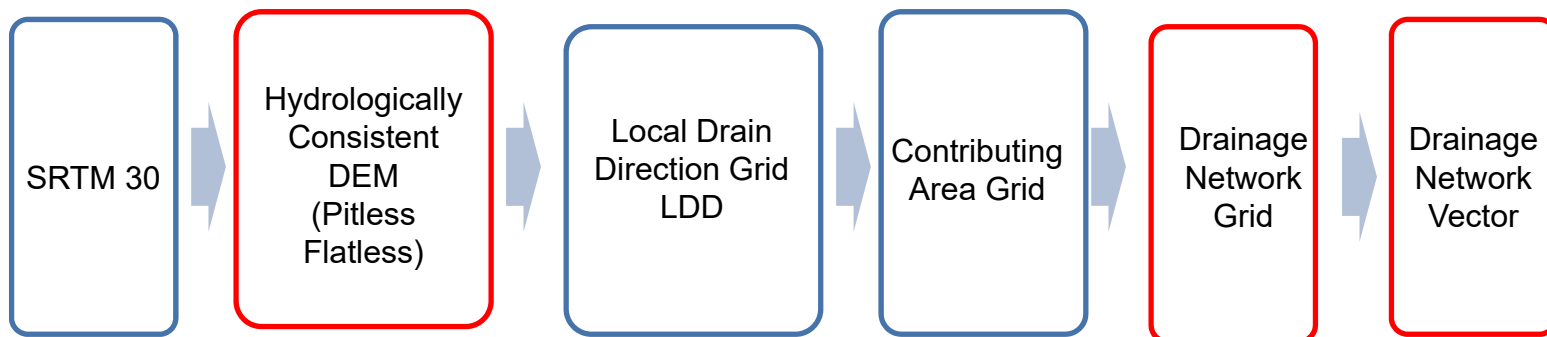
□ Dataset Extraction from DEM

- Drainage Network
- Basin Delineation
- Drainage Upscaling
- Flooding Area (HAND)



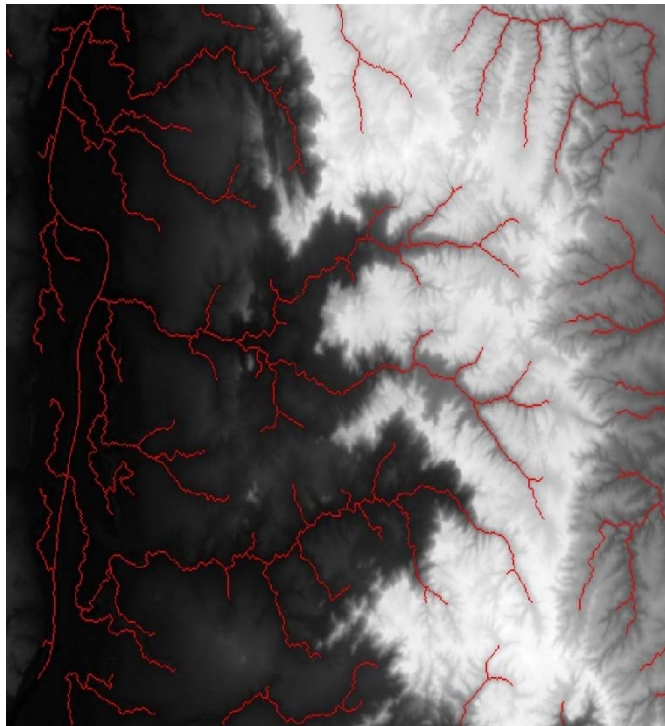
TERRAHIDRO DATA CREATION

- Drainage Network and Basin Delineation

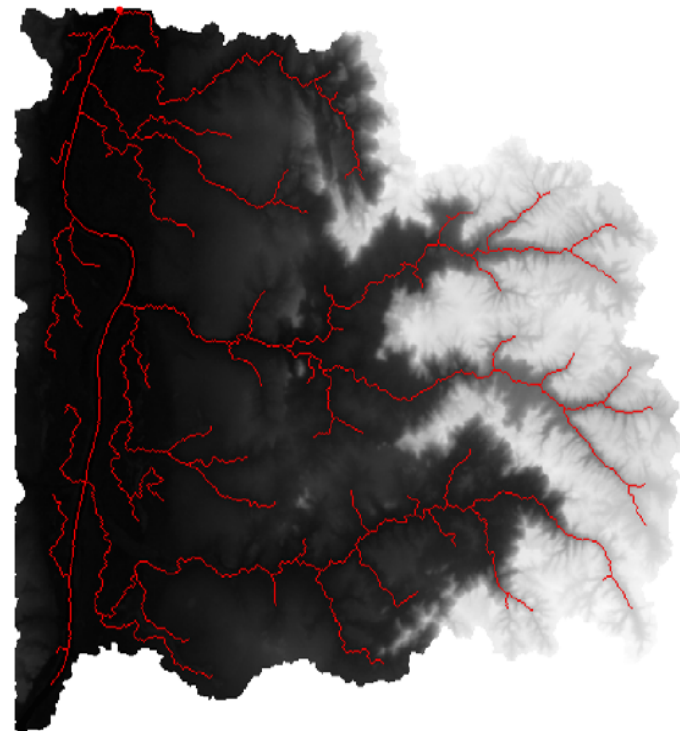


Data set produced in the drainage network pipeline

EXAMPLE DRAINAGE NETWORK AND BASIN



DRAINAGE NETWORK



WATERSHED DELIMITATION (BASIN)

TERRAHIDRO

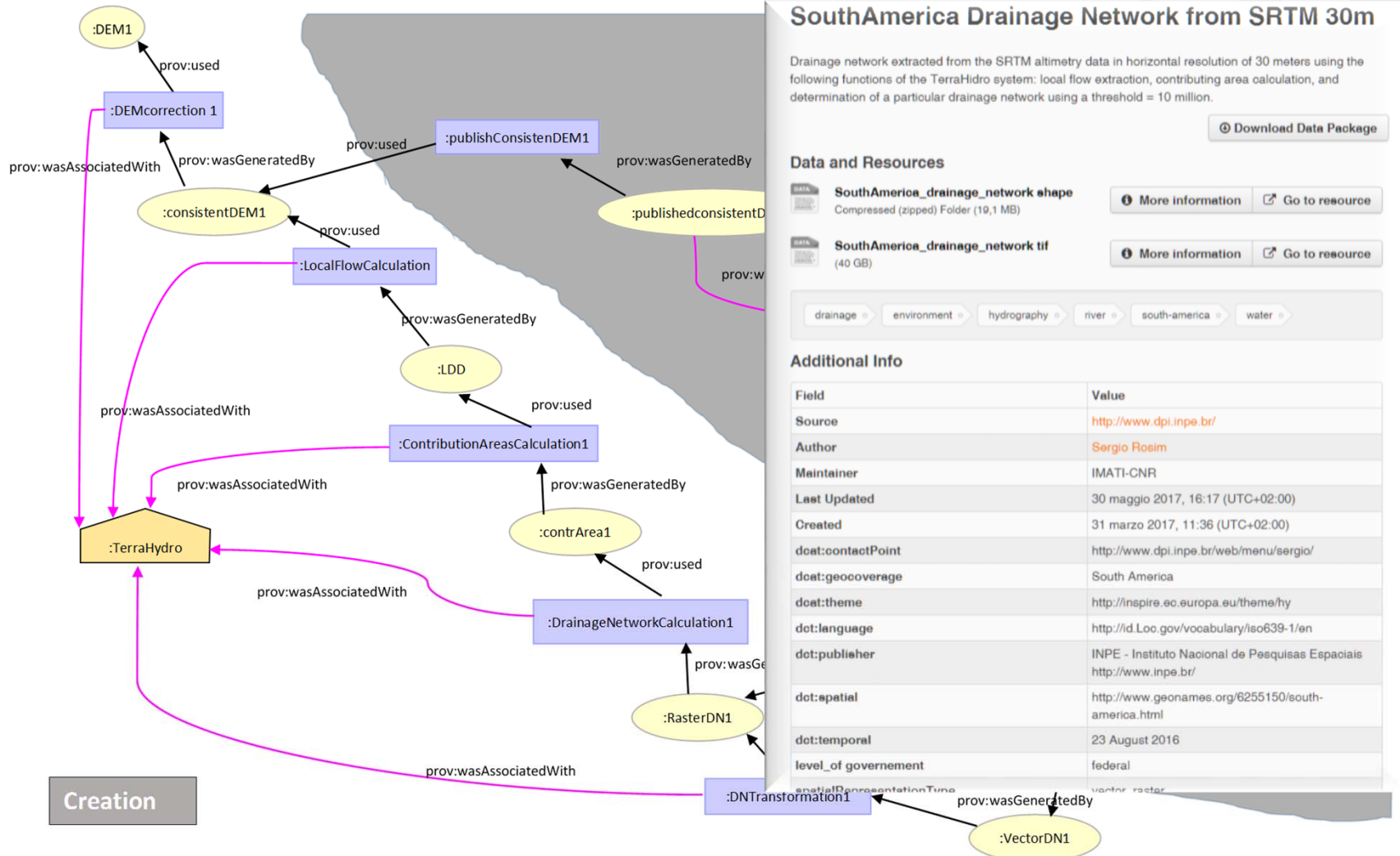
OPEN DATA PUBLICATION

Open Data Principle	Practice
Accessible	Rich set of metadata are provided to make data available to widespread audience – through a public point of access
Machine processable	Data are provided in machine readable formats to avoid technological barriers for third-party applications
Data are available in more than one format	Drainage networks and basins are supplied in more than one non-proprietary format (e.g. in vector format (shape), and raster format (.tif)). This allow reducing costs incurred in (re)processing data making available to a higher number of third-party tools.
Non-discriminatory	Data is available to anyone and downloadable from DataHub.io
License-free	Data are provided by the Creative Common license
Permanence	Data are stored in a centralized repository maintained by INPE
Usage Costs	Data are provided for free

DataHub portal

instance of the CKAN platform provided by the Open Knowledge Foundation

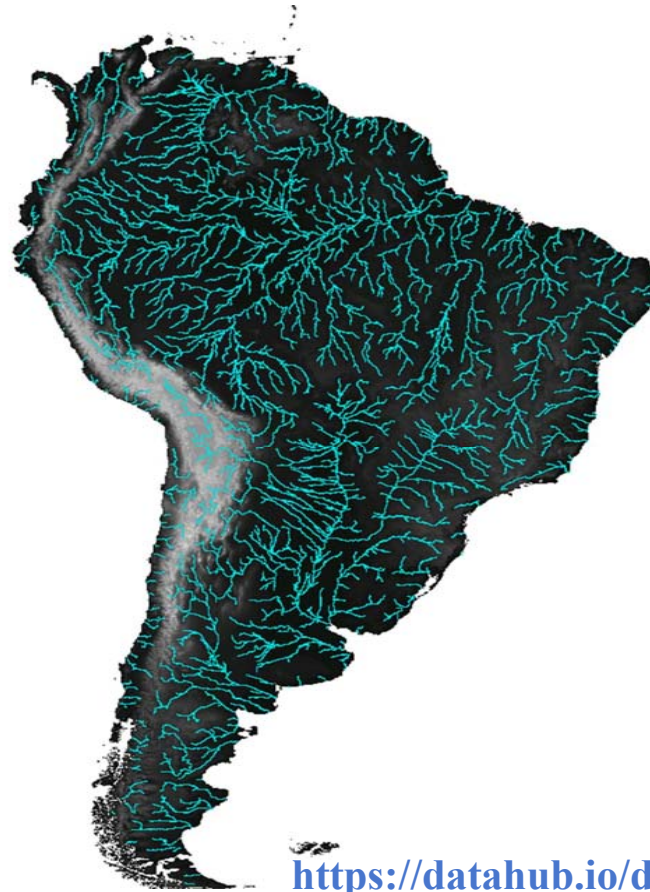
<http://datahub.io>



Provenance of Drainage Network creation and publication

(with W3C PROV)

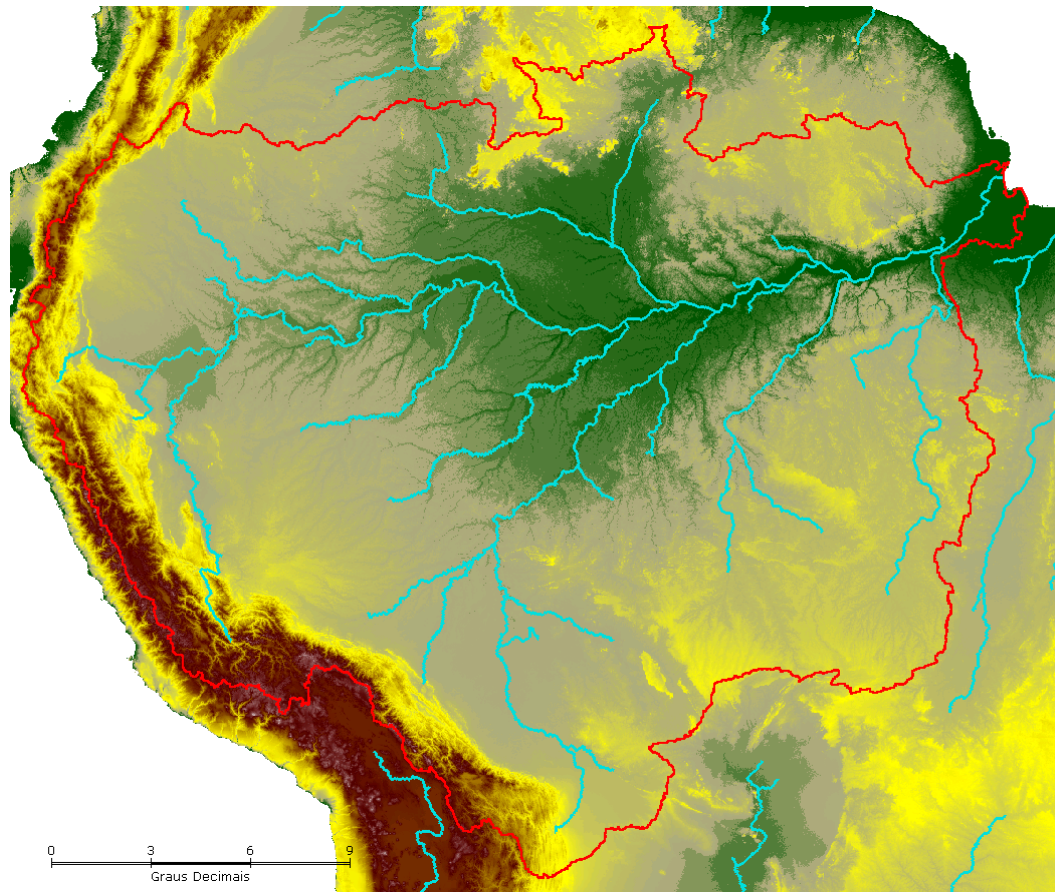
MONICA DE MARTINO
INSPIRE 2017



<https://datahub.io/dataset/southamericadrainage>

SOUTH AMERICA
DRAINAGE NETWORK

60.001 rows, 84,001 columns, **161.135.443 pits**



AMAZONIAN BASIN

32.400 ROWS, 38.400 COLUMNS, 65.670.466 PITS

Red: basin delimitation

Blue: drainages of main Rivers of Amazonian Basin

CONCLUSION

- **On Going Activity**
 - Publication of INPE drainage dataset as Open Data
 - South America drainages (SRTM 30m)
 - Amazon River basin
 - A description of the **WORKFLOW** of TerraHidro with PROV
- **NEXT**
 - Implementation of Workflow in RDF to make data reproducibility
 - Integration with other data (climate changing, agriculture)
 - Publication of further data (e.g drainage network of all world extracted d from SRTM 90 m)
 - Publication as Linked data

GRAZIE !
OBRIGADO !

Contact:

Monica De Martino
demartino@ge.imati.cnr.it
Alfonso Quarati, Riccardo Albertoni



Sergio Rosim
sergio.rosim@inpe.br

