

# Distributed Archiving Network

## An Archiving Protocol for the Future

Simon Margulies, Ivan Subotic, Lukas Rosenthaler

Imaging & Media Lab

University of Basel, Switzerland



### The Distarnet Protocol

#### What is it?

Distarnet is defined as an XML communication protocol and a set of rules for a distributed system.

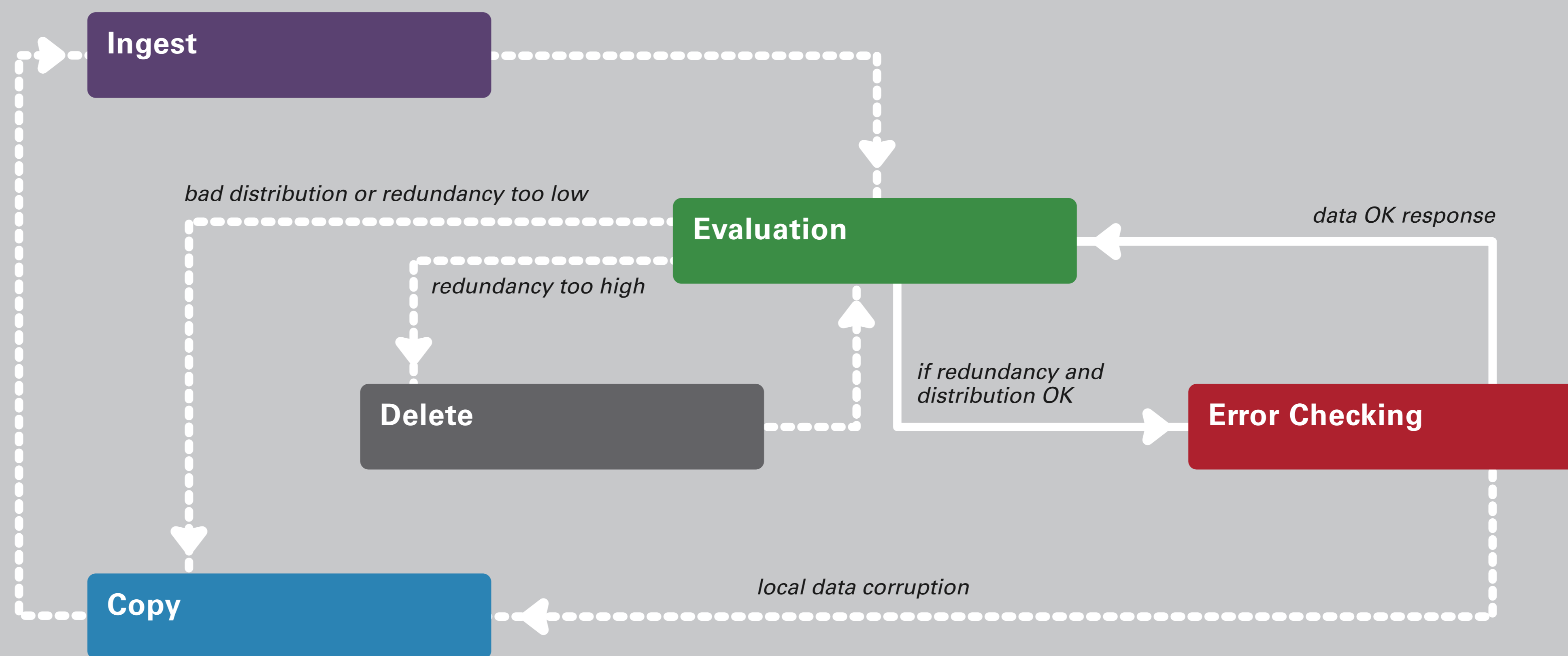
#### Key Features

P2P Architecture with strong encryption, controlled redundancy, fault tolerant recovery of network and data.

### Challenges

- + Massive growth in the production of digital data (digital-born or digitized). Data carriers unstable.
- + Digital data not readable by human eye. The need for a computer system to interpret and present.
- + Without description preserved, data becomes meaningless.

### Processes of Distarnet



### Evaluation

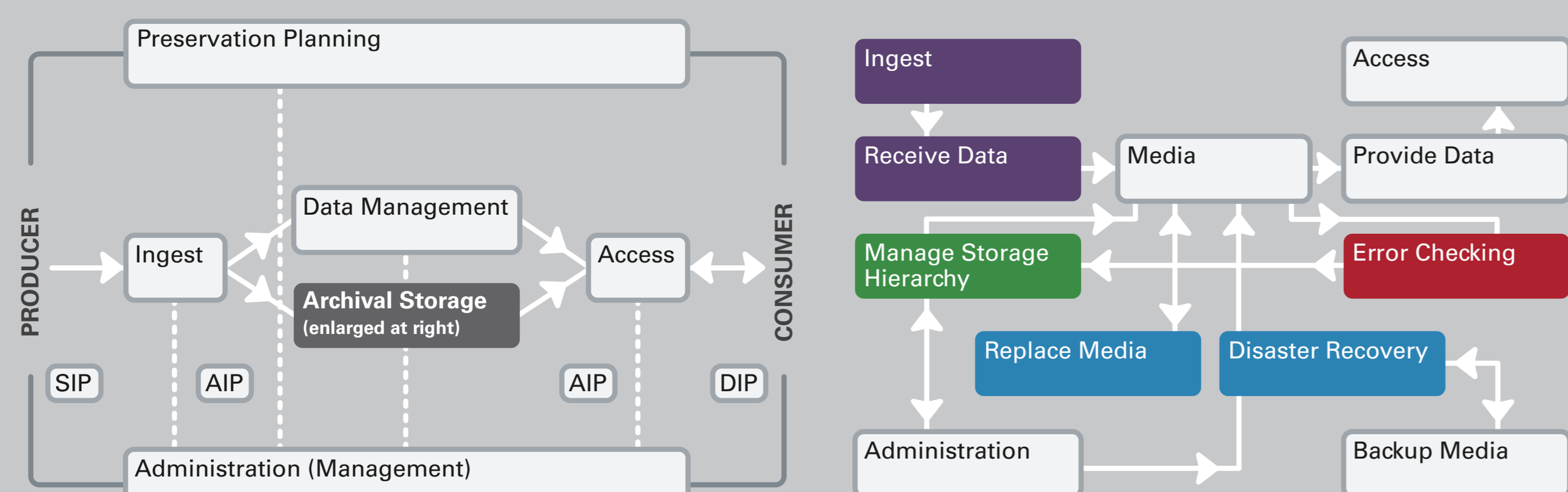
Analyzing the network to establish and balance distribution: finding best storage nodes according to network wide weighted (w) criteria (c), ingester node always wins evaluation:

- c<sub>1</sub> Geographical Location
- c<sub>2</sub> Space Ratio = Free / Total
- c<sub>3</sub> Uptime
- c<sub>4</sub> Speed
- w<sub>i</sub> 0 ≤ w ≤ 10

Nodes score points in every weighted criterion:

$$p_{node} = \sum_{i=1}^4 c_i * w_i$$

### OAIS vs. Distarnet Comparison

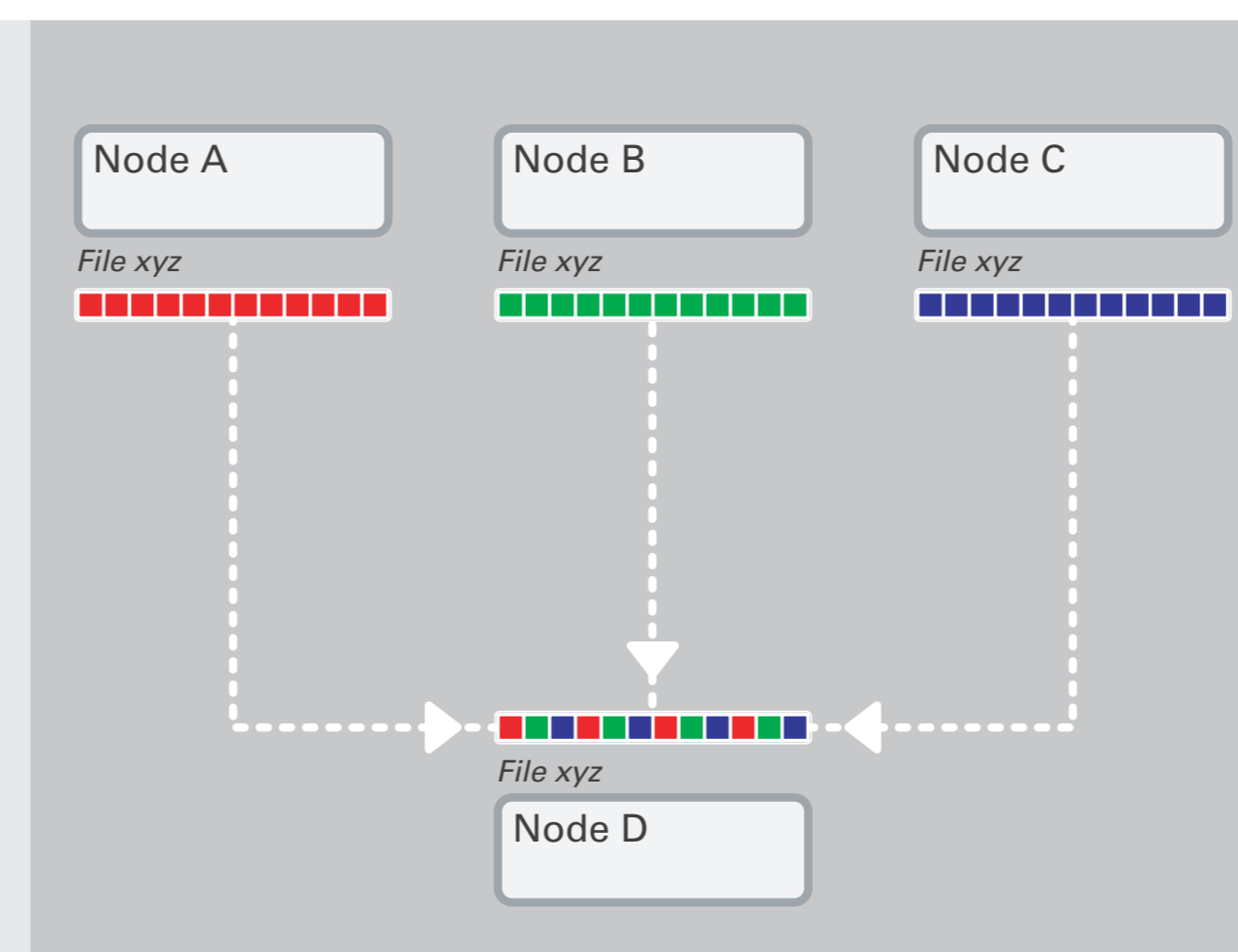


#### OAIS compared to Distarnet

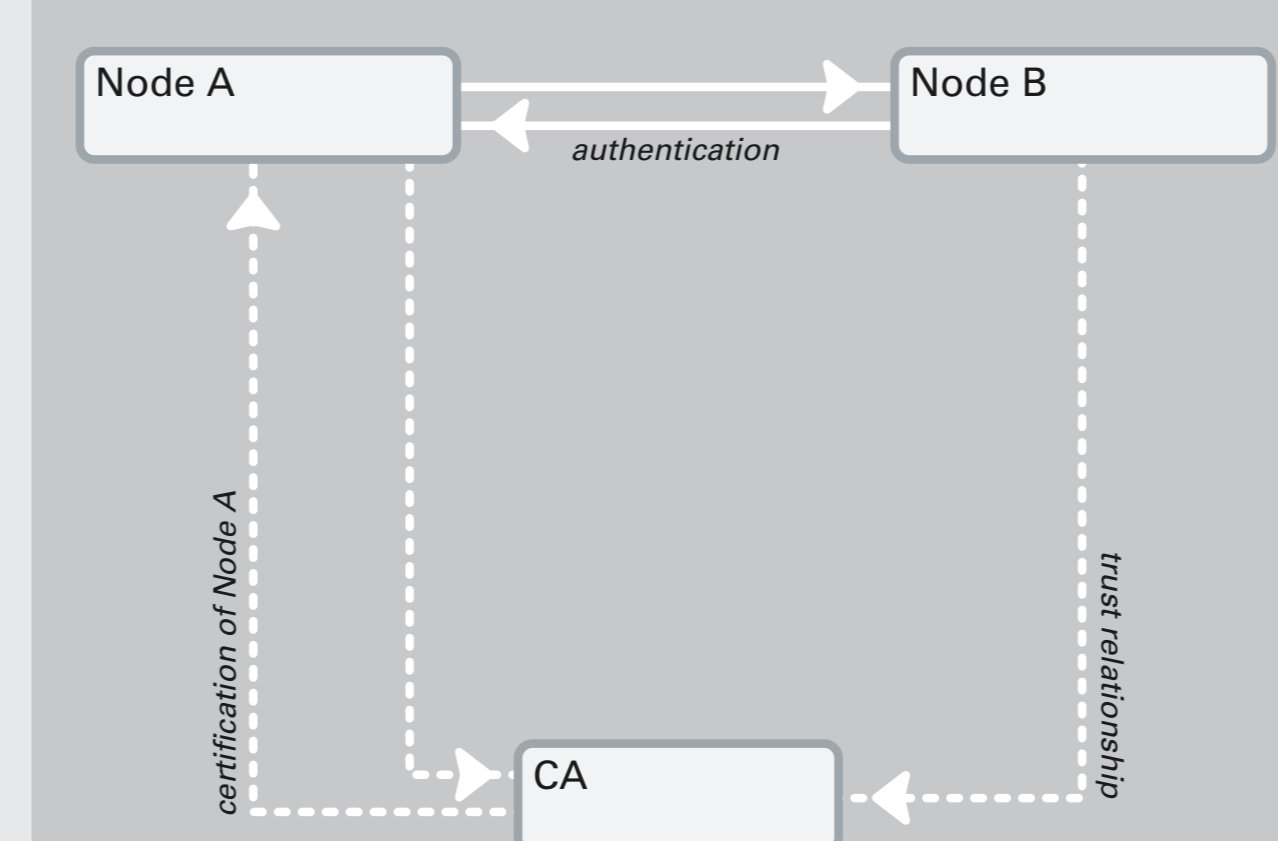
OAIS compared to Distarnet	dist.ar.net
Ingest/Receive Data	Ingest/Receive Data
Manage Storage Hierarchy	Evaluation
Error Checking	Error Checking
Replace Media	Copy
Disaster Recovery	Copy

### Distarnet Copy-Process / Security

- + Most crucial since it corresponds to the traditional data-carrier migration
- + Splitting files into chunks for optimal network performance
- + Secure Hash Algorithm used for consistency checks during the copy process

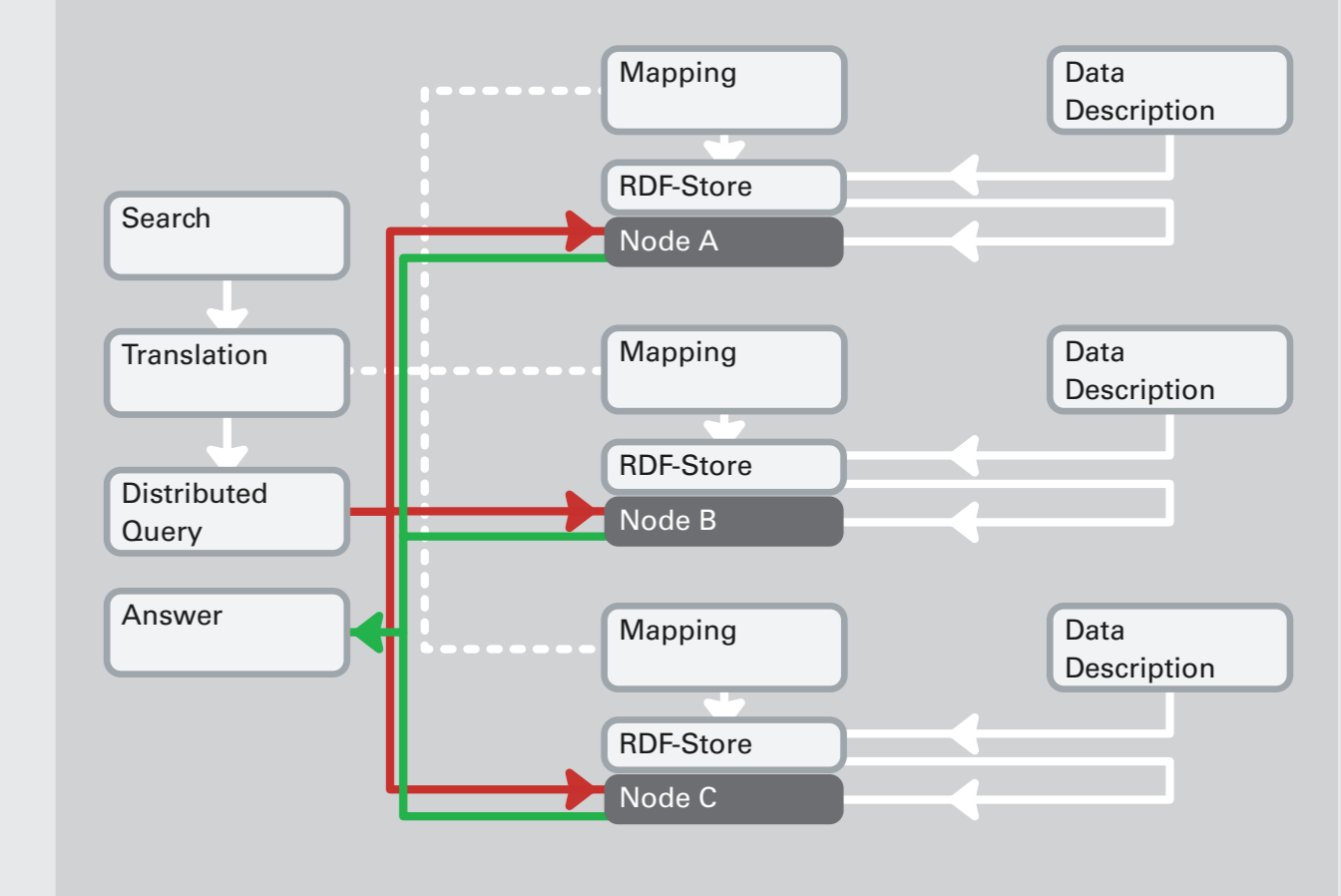


- + Closed user group in a P2P network
- + Authentication, integrity and non-repudiation in the communication of the peers over the internet
- + Public-Key Encryption (PKE) with a Public-Key Infrastructure (PKI)



### Metadata / Queries in Distarnet

- + Descriptive, administrative, structural and technical metadata in RDF
- + Semantical merging of queries to allow searching in diff. data descriptions



- + Routing queries through an overlay network based on the KADEMLIA lookup protocol
- + Network wide distributed hash table (DHT) for storing the search index
- + O(log N) messages

