

iMarine Tasks

Andrea Manzi

A blue wave graphic is located at the bottom left corner of the slide, partially overlapping a dark blue horizontal bar that runs across the bottom.

Main objectives of the project

Launch an Initiative aimed at establishing and operating a data Infrastructure supporting the principles of the

Ecosystem Approach to Fisheries Management and Conservation of Marine Living Resources

EA is a strategy for the integrated management of land, water and living resources that promotes conservation and sustainable use in an equitable way

- **Multi-disciplinary collaboration**
 - **Biodiversity:** inventories of biological information, conservation status, habitats,
 - **Fisheries:** socio-economic aspects of policy making, marine living resource assessment, fishing operations, processing and trade
- **Access and processing of a large amount of heterogeneous data for knowledge generation**
 - marine biology data, environmental data, catch data ,...
 - geospatial data, taxonomies, code-lists, ...

CERN Role: WP5 Operation

- T5.1: iMarine Data e-Infrastructure Operation
 - Operational procedures definition
 - Infrastructure incidents management
 - Infrastructure deployment and upgrade
 - establishment of links with the other infrastructures
 - WP3 Deployment Policies Implementation
- Upcoming Milestones and Deliverables
 - D5.1 iMarine Data e-Infrastructure Plan (Ready for Review) Nov 2011
 - Plans for infrastructure, resources tools and procedures.
 - Installation of GOCDDB for topology and Downtimes Management
 - MS25 Infrastructure nodes and gCube core services available, Jan 2012
 - “already” achieved cause we reuse the D4Science Infra

CERN Role: WP5 Operation

- T5.3: iMarine Data Infrastructure Availability, Monitoring and Accounting
 - monitor the status of the infrastructure and account its exploitation
 - verify the availability of the core functionality provided by the infrastructure
- Upcoming Milestones and Deliverables
 - MS27 Infrastructure availability, monitoring, and accounting tools deployed, Jan 2012
 - Using existing D4science tools
- Future Work (after April 2012)
 - Adoption of “standard” monitoring system, NAGIOS
 - Evaluation of the Nagios tools for remote services monitoring (NRPE, SSH) -> Alexs
 - Implementation of Storage Accounting

CERN Role: T9.2 The Data Transfer Facilities task

From DoW:

- “..provide a secure, reliable, and efficient solution for other deployed services to move different data types between remote infrastructure nodes”
- “..under different transfer protocols (e.g. srm-copy, gridFTP, HTTPS, BitTorrent, OPeNDAP, WCS, WMS, WFS, etc ..”
- “.. and the combination/optimization of state-of-the-art technologies (i.e. high-bandwidth networks, peer-to-peer)”.

CERN Role: T9.2 The Data Transfer Facilities task

- *“..data transfer mechanism to pass data by reference between infrastructure services by relying on a list of records that are part of a specific record set.”*
- *“..will support multiple transfer requests, minimize network load, not cause storage overload, manage transfer shares at service and user level, and allow data transfer parameterization.”*

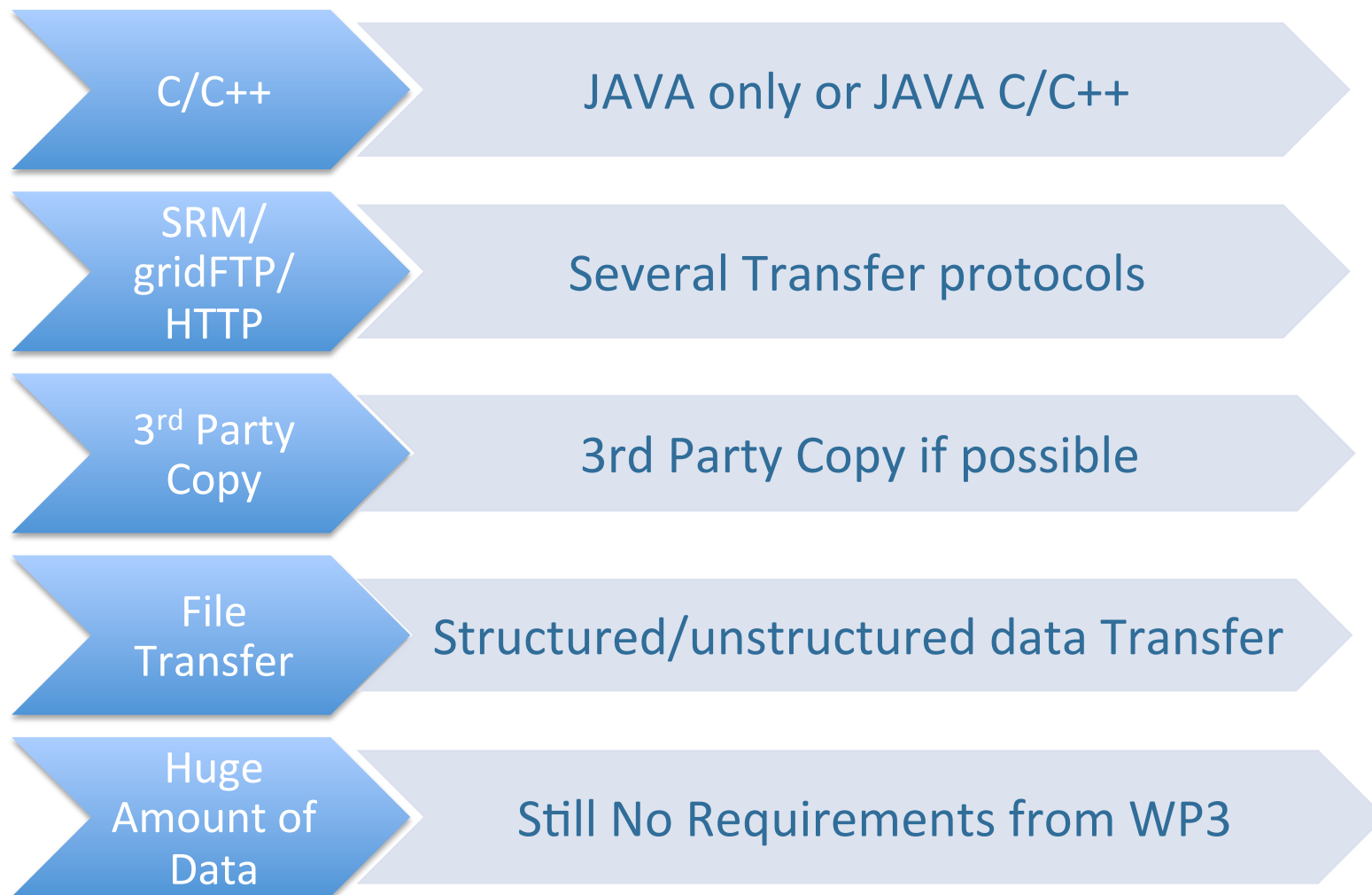
Data Access In gCube: Structured Data

- Content Manager Service (CM)
 - The service deals with edge-labelled and node-attributed trees, the gDoc trees.
 - New version based on Neo4j, NoSQL graph Database
 - gML defines an exchange format for documents (gCubeDocument) in which documents are constrained forms of edge-labelled trees:
 - the document: a self-contained unit of content within a collection of related units;
 - the metadata: a formal description of a document;
 - the annotation: a subjective observation or record about a document;
 - the part: a component of a document;
 - the alternative representation: a secondary manifestation of a document
 - CM Plugin Architecture mapping external Documents representations into gCUBEDocument -> Uniform Access to External and Internal Content.
 - CRUD ops, Streaming, Projections, Views, content URIs

Data Access In gCube: UnStructured Data

- Storage Manager Service (SM)
 - MongoDB using gridFS (file chunking)
 - RESTFul interface
 - OCCI compliant (standard from OGF)
 - Open Cloud Computing Interface, Protocol and API for all kinds of Management tasks
 - Distribution implemented by MongoDB
 - Not in production yet
- DPM, still no “real” integration plans
 - Currently exploited by OCR-ring using wrappers around lcg-util
 - GFAL 2 and HTTP interface evaluation
 - GFAL 2 JAVA API ?
 - Discussion with Adrien

FTS → gCube Data Transfer



gCube Data Transfer Agent

Need for a solution to enable:

- Different data transfer protocol
- Possibly Integrated with FTS 3 to:
 - Enable sharing development with EMI team
 - Push gCube extensions for FTS to be deployed on EGI infra



gCube Data Transfer Agent

- Deployed on each infrastructure node
- Plugin architecture
- WS interface

FTS 3 integration in gCube ?

- gDTA design:
 - Plugins implementation for transfer protocols
 - gridFTP , BitTorrent, etc
 - 2 Access interfaces (CM and SM)
- Possible FTS core extensions:
 - extensions to handle Content URIs, sms:// , cms://.
 - gDTA cannot connect to FTS DB
 - FTS pushes jobs to gDTA
- Lots of Open points:
 - Security
 - FTS3 time plan
 - Missing Requirements
- Collaboration with Zsolt and Michail

Plan till April 2012

- The first milestone:
 - MS38 Data Transfer Facilities Specification (April 2012)
“The specification of the facilities for data transfer are published in dedicated wiki pages”
- Strict depends on :
 - MS37 Data Access and Storage Facilities specification (Feb 2012)

Feasibility study on FTS Integration (till end of Jan)

First data transfer agent design (Feb-Mar)

Complete gCube Data Transfer specification and first implementation(Apr)



D9.1 iMarine Data Management Software (Apr)

Thanks!