AENEAS WP4

Analysis of Global SKA Data Transport and Optimal European Storage Topologies

WP4 Objectives

- Investigate and demonstrate the data transfer and storage techniques required to confirm the viability of a distributed computing and network architecture for a European Science Data Centre.
- Collaborate with South Africa and Australia to address the challenges of moving large data volumes. PoC trials and support the work of WP3
- Moving up the stack, Study of: data access protocols, data transfer protocols, replica and transfer management, data moving applications
- Output: design and best practice recommendations for the construction of the network infrastructure, data transfer and storage required for an ESDC together with cost models for European and Global connectivity.

WP4 Partners & Stakeholders

Work package number	4		Lead beneficiary		GEANT LTD
Participant number	2	4	5	6	8
Short name of participant	UMAN	INAF	Chalmers	GEANT Ltd	Jülich
Person/months per participant:	6	1	20	22	9
Start month	1		End month	36	

Stakeholders: AARNet, ASTRON, CSIRO, IT,
JIV-ERIC, MPI Bonn, SANReN, STFC

WP4 Tasks

- Task 4.1: Evaluation of existing data transfer protocols, storage sub-systems and applications

  Start →T0 month

- Task 4.2: Inventory of the storage and network capabilities of existing and planned European Facilities for SKA

  Start → T0+6 months

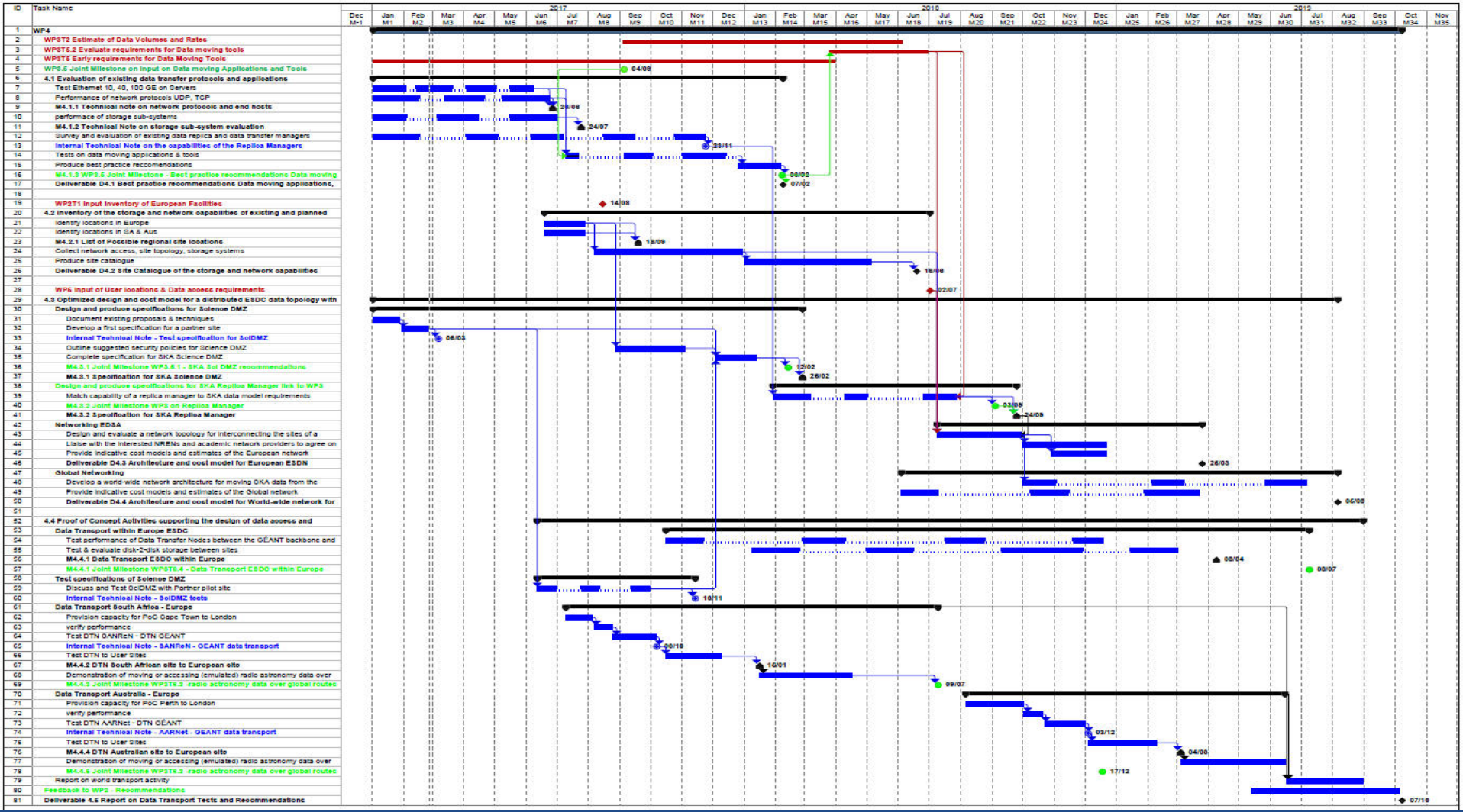
- Task 4.3: Optimized design and cost model for a distributed ESDC data topology with world connectivity

  Start → T0 & T0+8 months

- Task 4.4: Proof of Concept Activities supporting the design of data access and transport within Europe and from the Host countries to Europe

  Start → T0+6 & T0+19 months

WP4: Task Gantt Chart



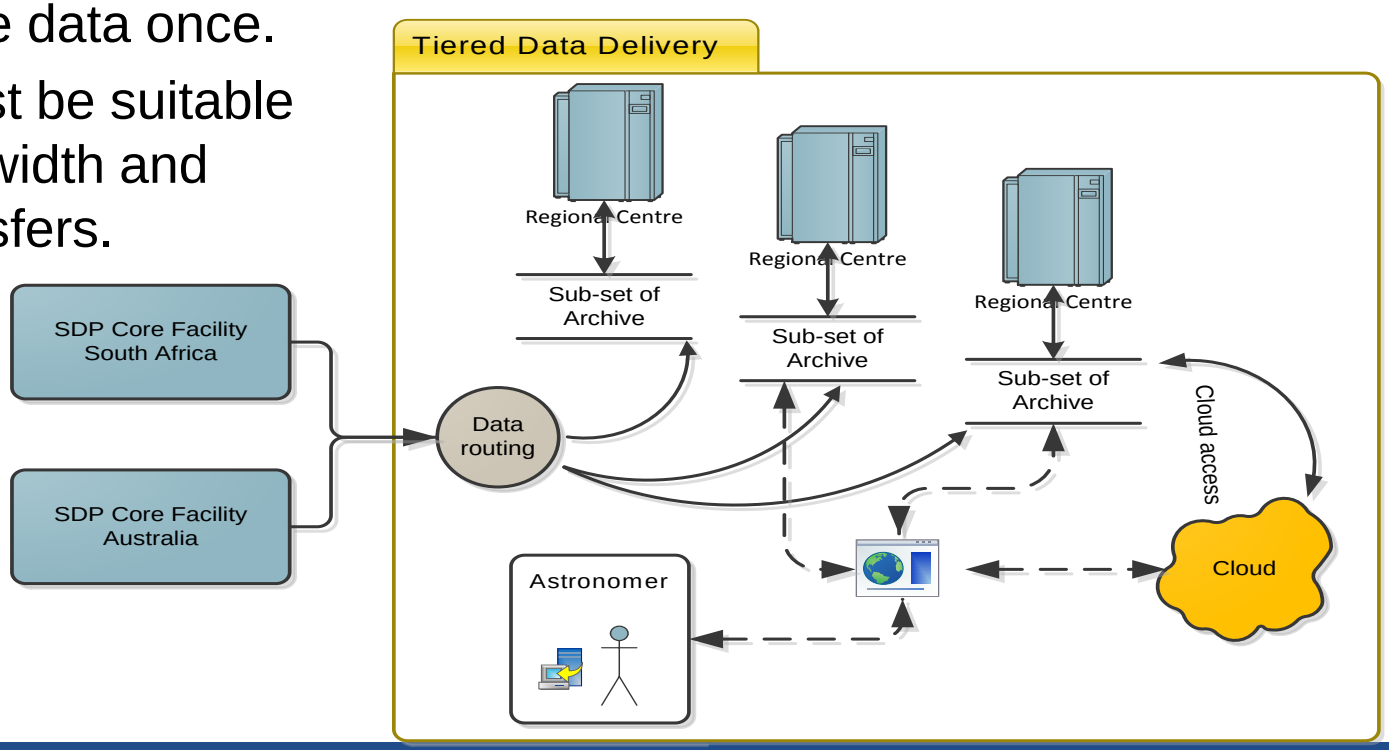
Task 4.1: Evaluation of existing data transfer protocols, storage sub-systems and applications

Partners: Chalmers (lead), GÉANT Ltd, Jülich, INAF, UMAN Stakeholders: CSIRO, SANReN, IT

- Test network protocols and end-host performance
  - 10 40 100 GE, UDP, TCP & variants, RDMA
  - The effect of CPU socket, core and app locations
- Test & evaluate storage sub-systems and file systems
  - i/o performance & scaling
  - Data curation and safety of the data – modern RAIDx !
- Evaluate & test data transfer protocols & applications
  - High performance, long distance environment
  - Gridftp, aspera, http, …
- Survey and evaluate data replica and data transfer managers
  - dCache, Hadoop, Ceph, xroot, …

Delivering the Data – Requirements Input

- SDP HPC processing places SKA data in telescope archives.
- Tiered Model to provide access to the astronomy community:
  - Having a replica is a basic requirement.
  - Only move the data once.
  - Protocols must be suitable for high bandwidth and real-time transfers.
- WP3 & WP5



Task 4.2: Inventory of the storage and network capabilities of existing and planned European Facilities for SKA

Partners: INAF (lead), GÉANT Ltd Stakeholders: ASTRON, IT

- Online catalogue of technical capabilities
  - Connectivity of the site
  - Network topology to compute and storage elements
  - Details of the storage technology
- WP 2.1 provides an initial list of possible ESDC sites including
  - Existing & proposed radio astronomy locations
  - National Tier 1 centres

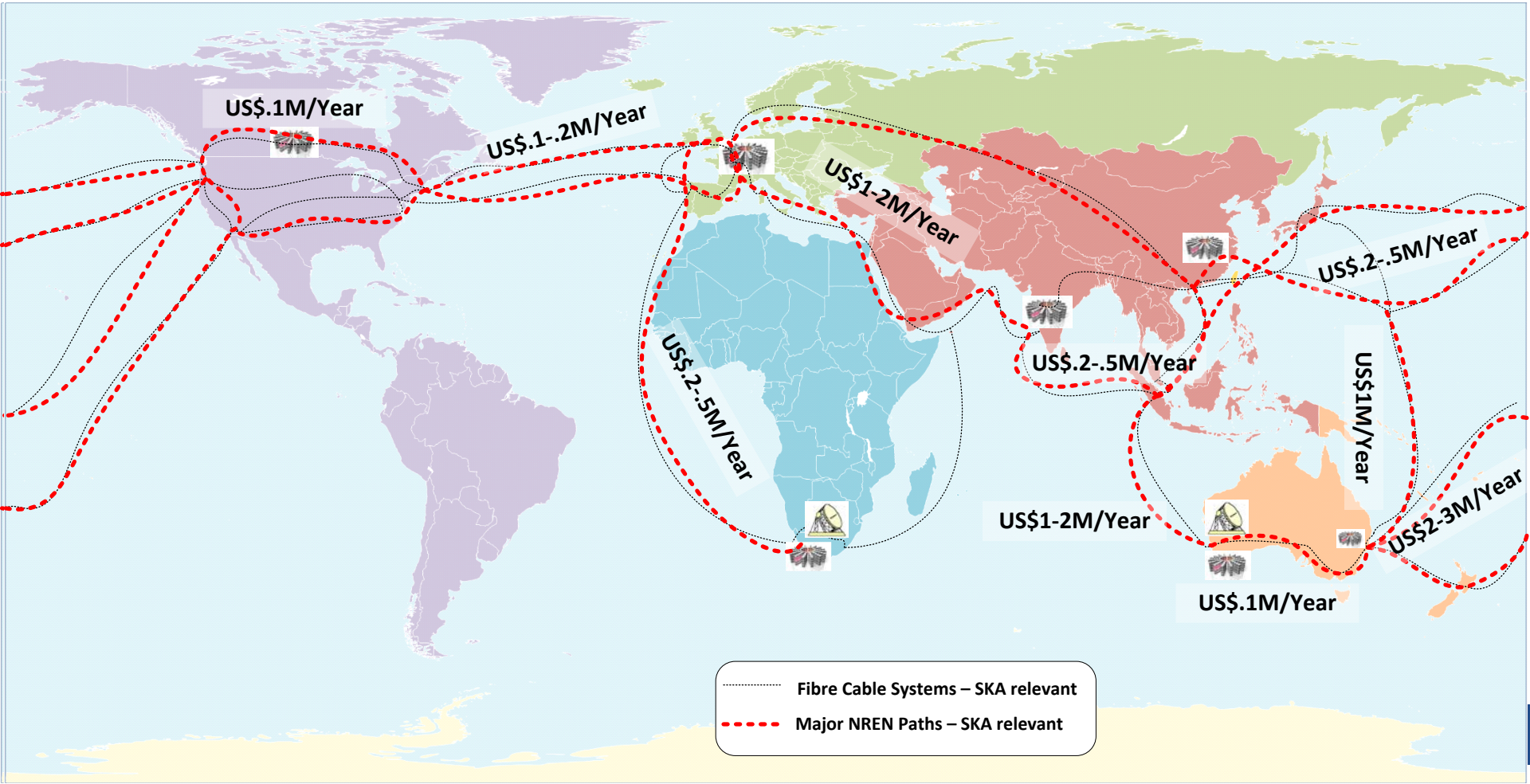
Task 4.3: Optimized design and cost model for a distributed ESDC data topology with world connectivity

Partners: GÉANT Ltd (lead), INAF, Chalmers, Jülich, UMAN Stakeholders: AARNet, CSIRO, SANReN, IT

- Close links with WP3 & WP5 establish network requirements
  - General user access
  - High bandwidth data moving
- Design & spec. network infrastructure for a distributed ESDC
  - Support internal ESDC, remote access & data transfers with DTNs
  - What form of connectivity?
  - Formation of De-Militarised Zones at the sites
- Assess capability & use of Software Defined Networking
- Design & specify a SKA replica manager for ESDC
- Develop a global network architecture for moving SKA data
- Provide indicative cost models for European and global links.

Possible Global Network Paths

- 10 year IRU per 100Gbit circuit 2020-2030
- Guesstimate of Regional Centres – based on SKA member locations
- **Some funding model will be needed for SKA.**

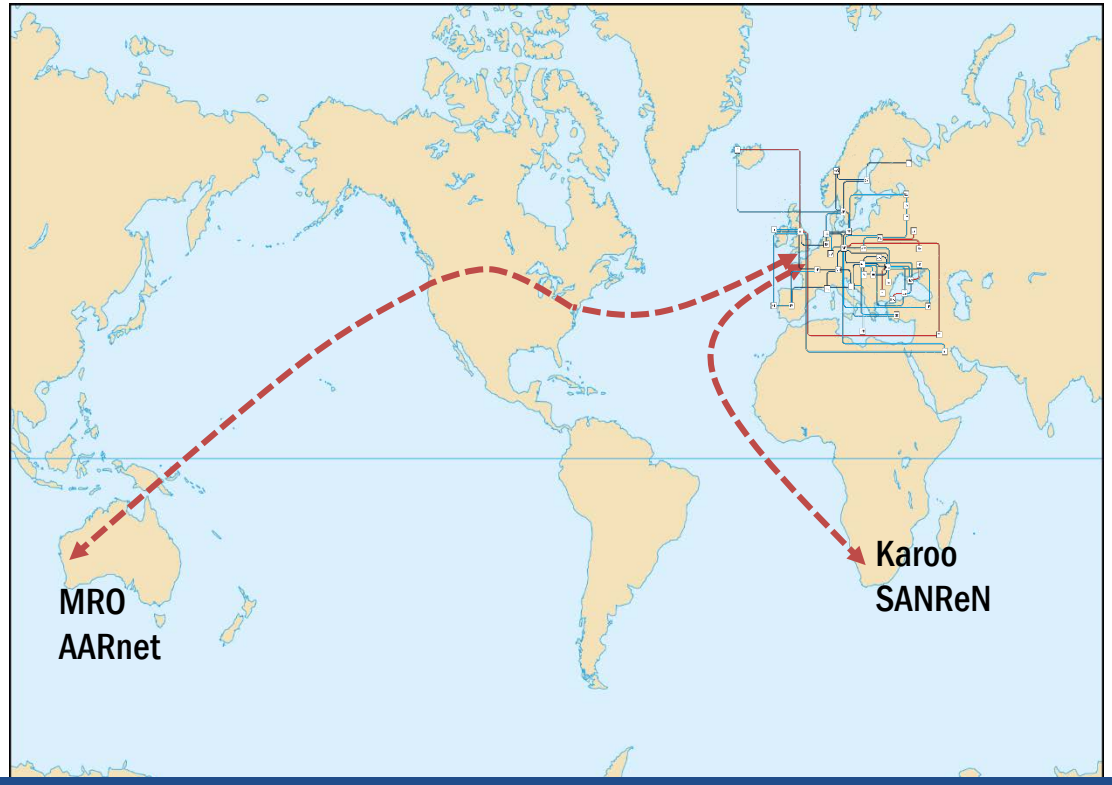


Task 4.4: Proof of Concept Activities: Data access and transport within Europe and from the Host countries to Europe

Partners: GÉANT Ltd (lead), Chalmers, UCAM, UMAN Stakeholders: AARNet, CSIRO, JIV-ERIC, SANReN, IT

- Work with End Sites & GEANT/NRENs to set up European paths
- Work with AARNet & SANReN to set up inter-continental paths
- Test DMZ specs as a pilot at an AENEAS partner site
  - Facilitate discussions of (existing) security policies
  - Testing performance
- Measure performance between GEANT DTN & likely ESDC sites
  - Network tests
  - Storage – storage transfer tests
- With AARNet & SANReN test inter-continental performance
  - Network tests – protocols & long haul effects – multiple 10 Gigabit
  - Storage – storage transfer tests (NREN-NREN & EndSite-EndSite)
  - Move Radio Astronomy data with WP3 (EndSite-EndSite)

- With AARNet & SANReN test inter-continental performance
  - Network tests – protocols & long haul effects – multiple 10 Gigabit
  - Sustained data transfers
  - Storage – storage transfer tests (NREN-NREN & EndSite-EndSite)
  - Move Radio Astronomy data with WP3 (EndSite-EndSite)



WP4 Deliverables

- D4.1 Best practice recommendations Data moving applications, protocols and storage

  → T0+14 months

- D4.2 Site Catalogue of the storage and network capabilities

  → T0+18 months

- D4.3 Architecture and cost model for European ESDN network

  → T0+27 months

- D4.4 Architecture and cost model for World-wide network for SKA

  → T0+32 months

- D4.5 Report on Data Transport Tests and Recommendations

  → T0+34 months

Interactions and Dependencies with other WP

- Strong supporting links between WP3 and WP4
  T4.1 & T3.2 Data volumes & rates
  T4.1 & T3.5.2 requirements for data moving tools
  T4.4 & WP3 PoC tests of science data
- Many joint Mile Stones

Input

- From WP2: T1 to form list of possible sites for ESDC. T4.2 (M9 Jul17)
- From WP5: User locations & Data access requirements T4.3 ~M20 Jul18

Joint Milestone
Related WP
WP4 Lead

Milestone number	Milestone name	Related work package(s)	Due date (in month)	Means of verification
6	Protocols and end hosts evaluation	WP4	7	Technical note written
7	Storage sub-systems evaluation	WP4	8	Technical note written
10	Joint Milestone (WP4) on data moving applications & tools	WP3 WP4	9	Internal memo
11	List of possible regional site locations	WP2 WP4	9	List of possible sites established
19	Data transfer test South African site to European site	WP4	13	Technical note written
20	Joint Milestone (WP4) on SKA Sci DMZ recommendations	WP3 WP4	14	Internal memo
21	Best practice recommendations Data moving applications, protocols and storage	WP3 WP4	14	D 4.1 written
22	Specification for SKA Science DMZ	WP3 WP4	14	Specification document written
25	radio astronomy data over global routes from South Africa to Europe	WP3 WP4	18	WP3 Technical note written
27	Joint Milestone (WP4) on demonstration of moving data from observatory sites (SA) to ESDC	WP3 WP4	19	Demonstration completed
30	Joint Milestone (WP4) on data replica manager	WP3 WP4	21	Internal memo
31	Specifications for SKA Replica Manager	WP3 WP4	21	Specification document written
33	Joint Milestone (WP4) on demonstration of moving data from observatory sites (AUS) to ESDC	WP3 WP4	24	Demonstration completed
35	Data transfer test Australian site to European site	WP4	27	Technical note written
36	Report on Data Transport ESDC within Europe	WP3 WP4	28	Technical note written
37	radio astronomy data over global routes from Australia to Europe	WP3 WP4	30	WP3 Technical note written
40	Joint Milestone (WP4) on demonstration of moving data within ESDC	WP3 WP4	31	Demonstration completed

WP4 Moving Forward

T4.1:

- Locate test systems and login access. Started investigations of protocols and end systems.
- Survey and evaluate data replica and data transfer managers, file systems and storage sub-systems.

T4.2 :

- Consideration of technology required for the online catalogue.
- Location of the server – www.aeneas2020.eu

T4.3 :

- Initial spec for a DMZ on a SKA site

T4.4 :

- PoC tests with Australia and South Africa
- Establish locations and suitable servers in the NRENs
- Formed contacts with other projects: eg Enlighten my Research (NZ, Aus, NL); Asterics

Questions ?

Thanks to Richard Hughes-Jones