Applying Risk Management to Support SLA Provisioning

G. Birkenheuer[1], D. Battré[2], M. Hovestadt[2], O. Kao[2], and K. Voss[1]

[1] Paderborn Center for Parallel Computing, Universität Paderborn

[2] Technische Universität Berlin

Germany

Motivation

- Best Effort is not Enough

- Grid commercialisation requires establishment of Service Level Agreements (SLAs)
- Agreeing an SLA is a business risk for a provider
- risk management processes are required

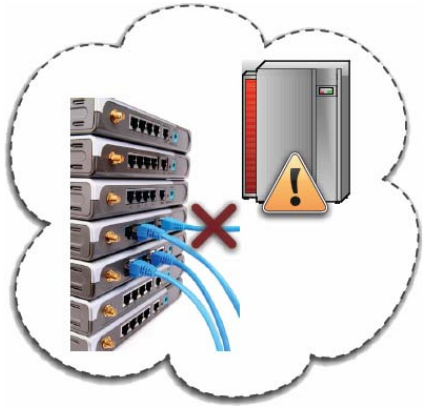
Agenda

- Why Risk Management in Grids?
- Grid Risk Management (RM) Process
  - FERMA Standard
  - implementing FERMA to Providers RM
- Conclusion

Why Risk Management?

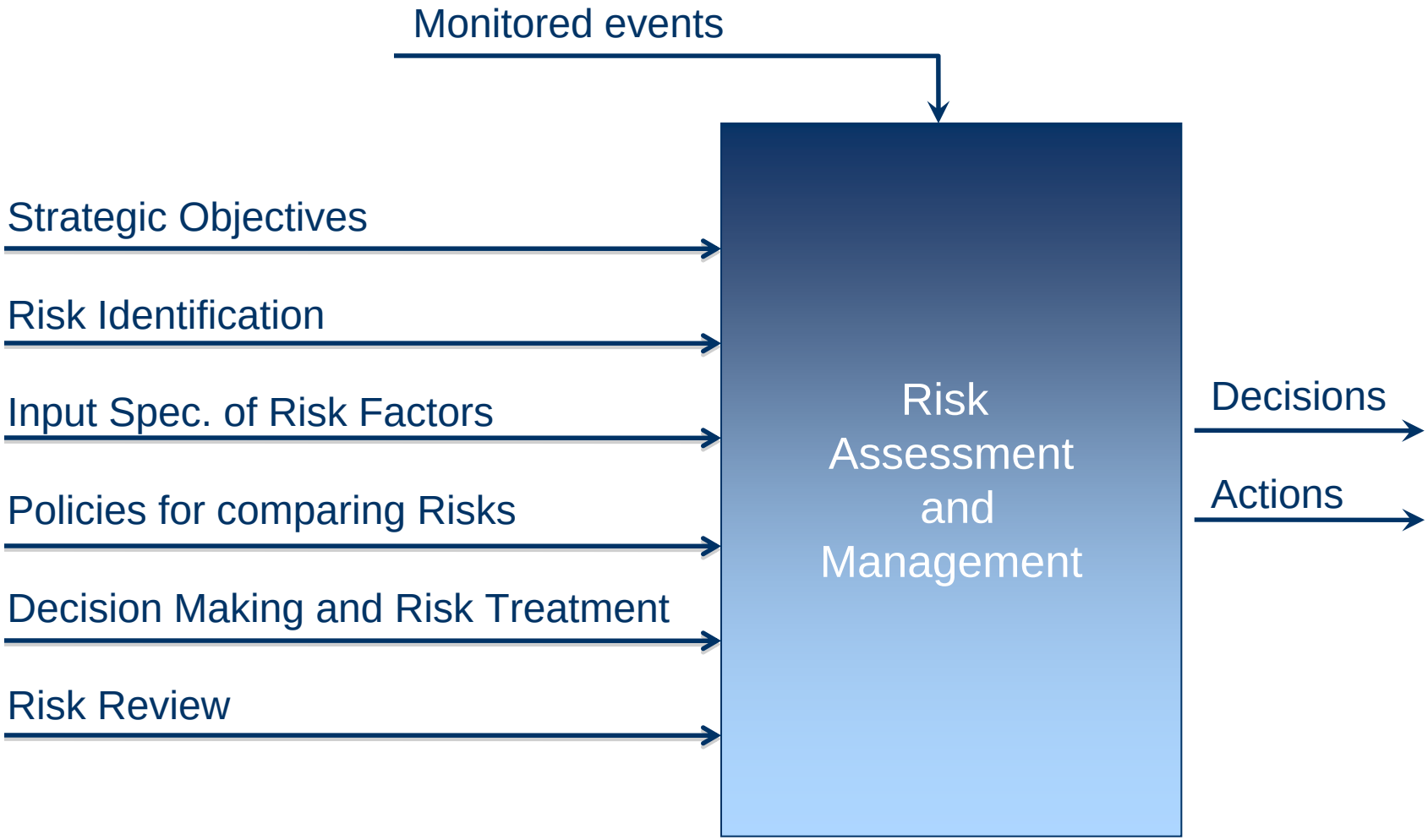
- Job failure rates in Grids are high
  - TeraGrid: 10 - 45 % [Khali 06]
  - Grid3: 27% - even with 5-10 retries [Dumi 05]

- Reason is frequency of resource outages
  - Grid'5000: MTBF ~14 hours [Iosu 05]



- Consequences:
  - providers are reluctant to accept strict SLA requests
  - users doubt that an SLA will be fulfilled

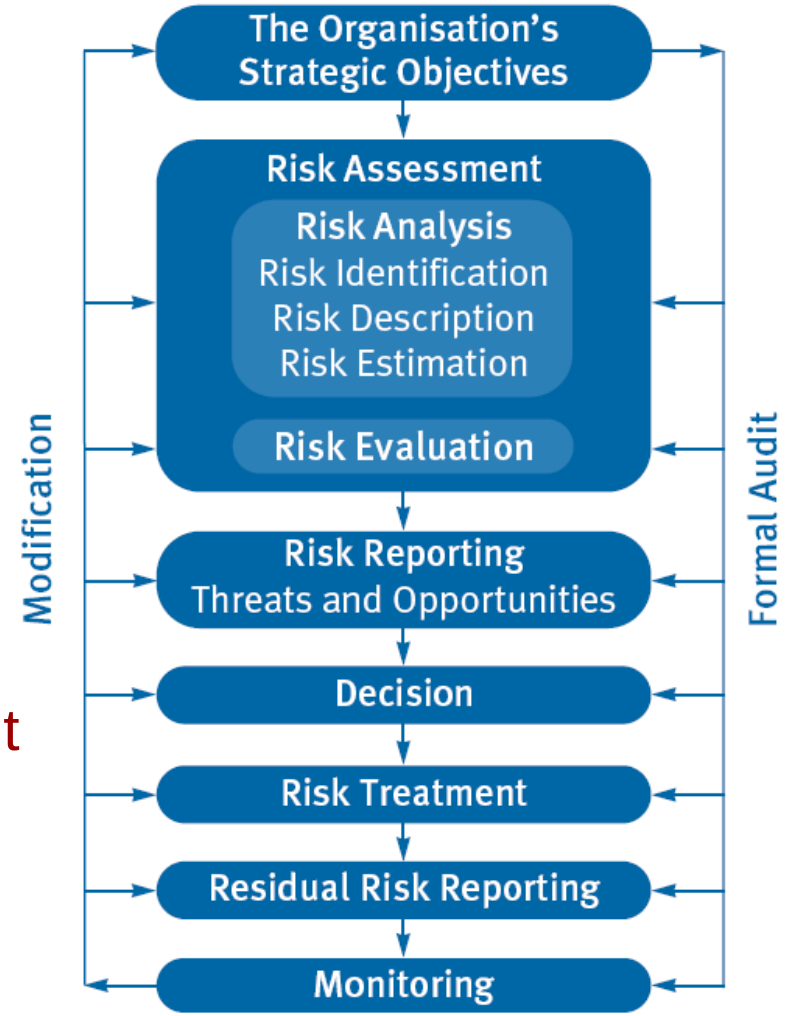
Risk aware Grid Fabric Layer



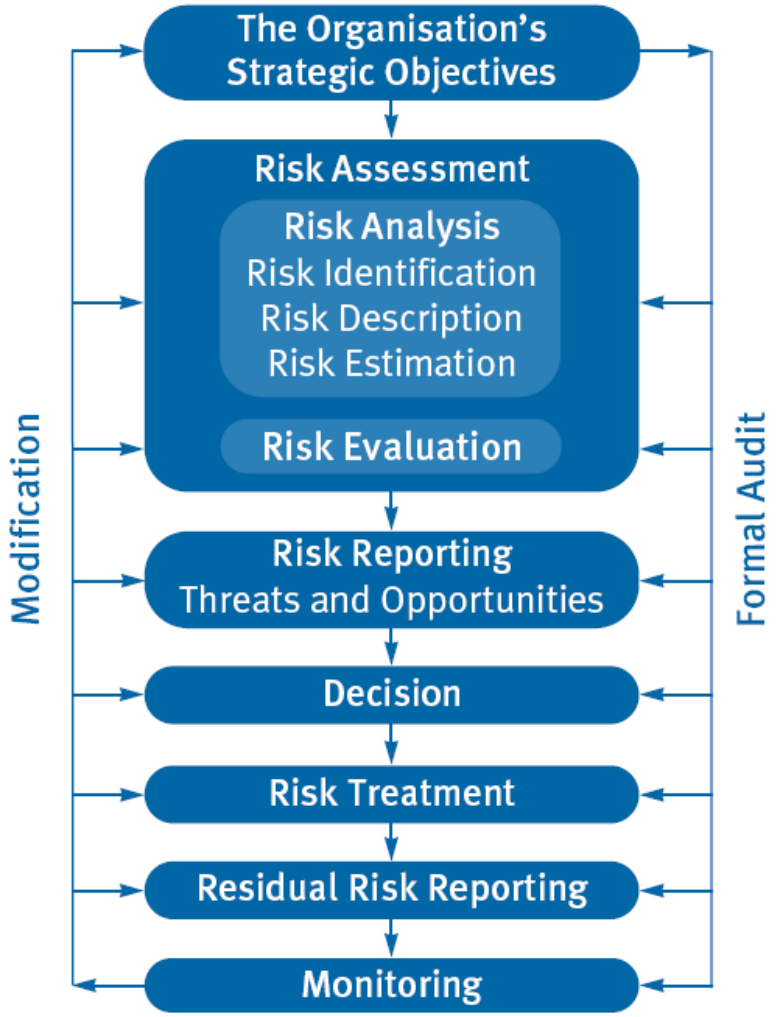
Are there any standards?

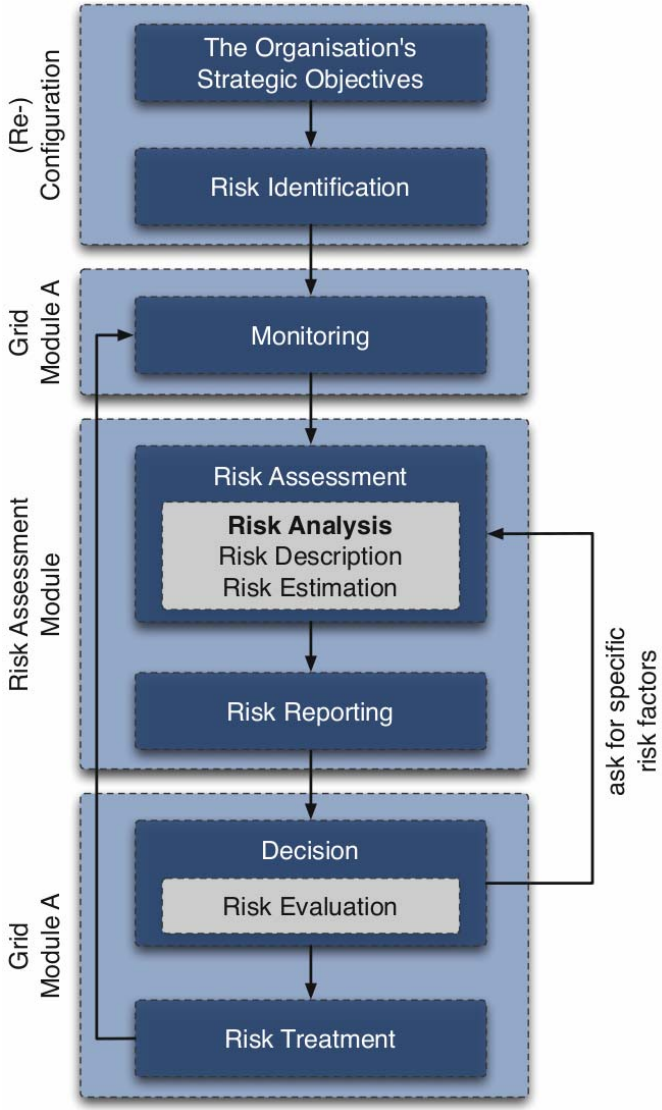
FERMA Standard

- Developed for any RM plans
- Designed for any kind of enterprise
- Manual procedure
  - Grid does run automatically
- Often applied once
  - Grid needs steady adjustment



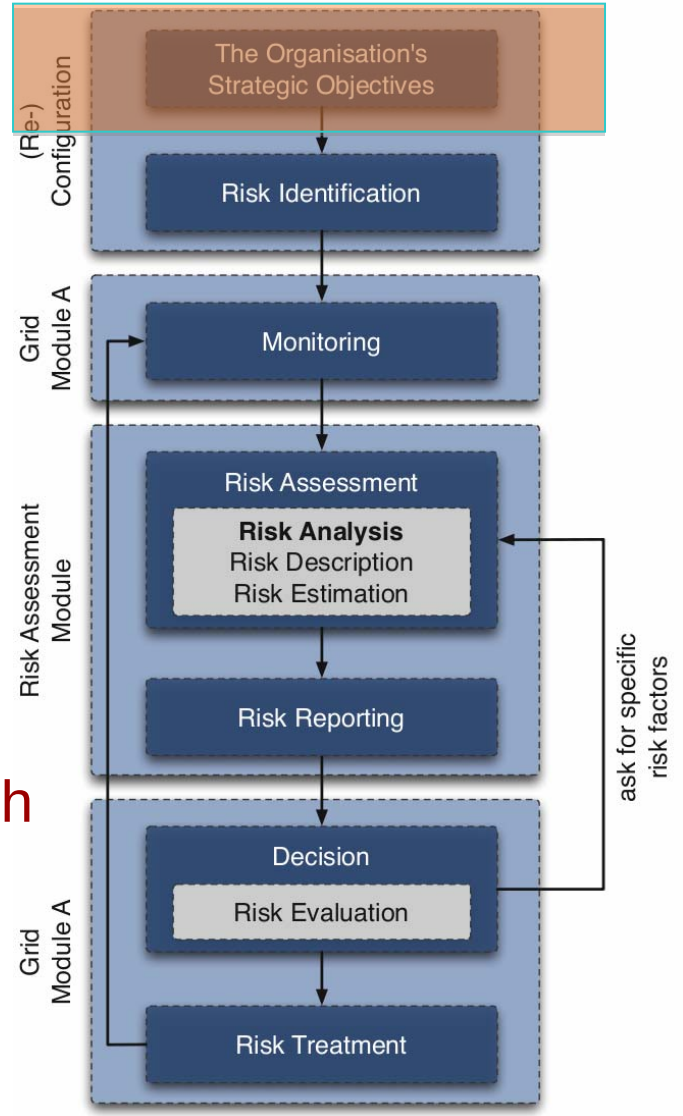
Implementing FERMA to Providers RM





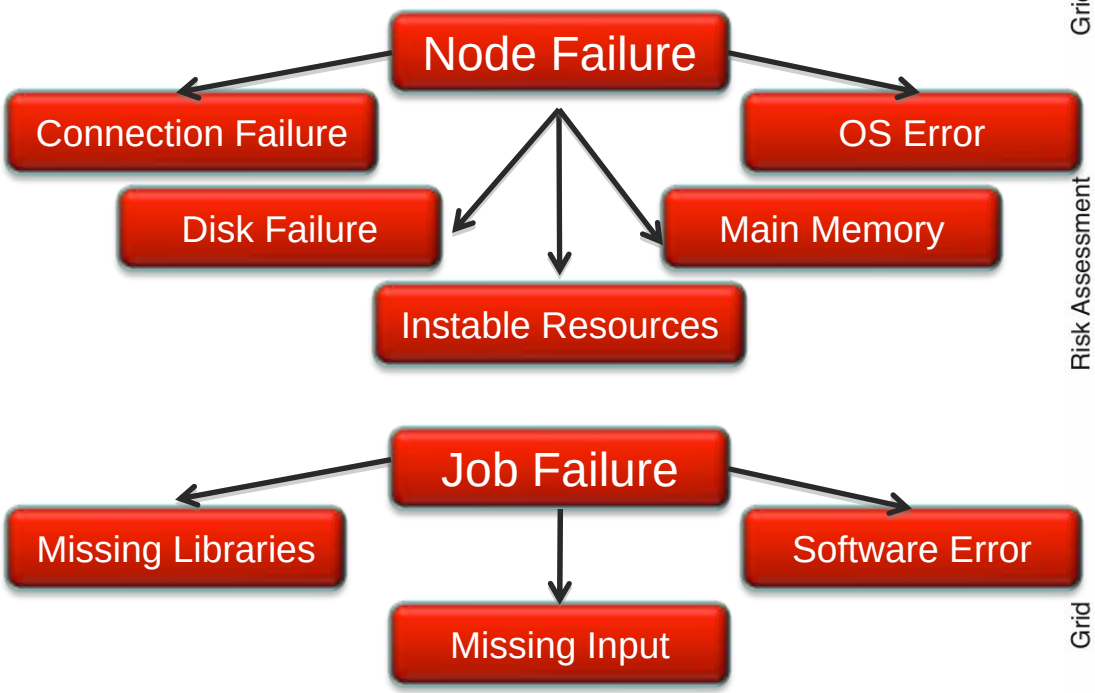
Specification of Strategic Objectives

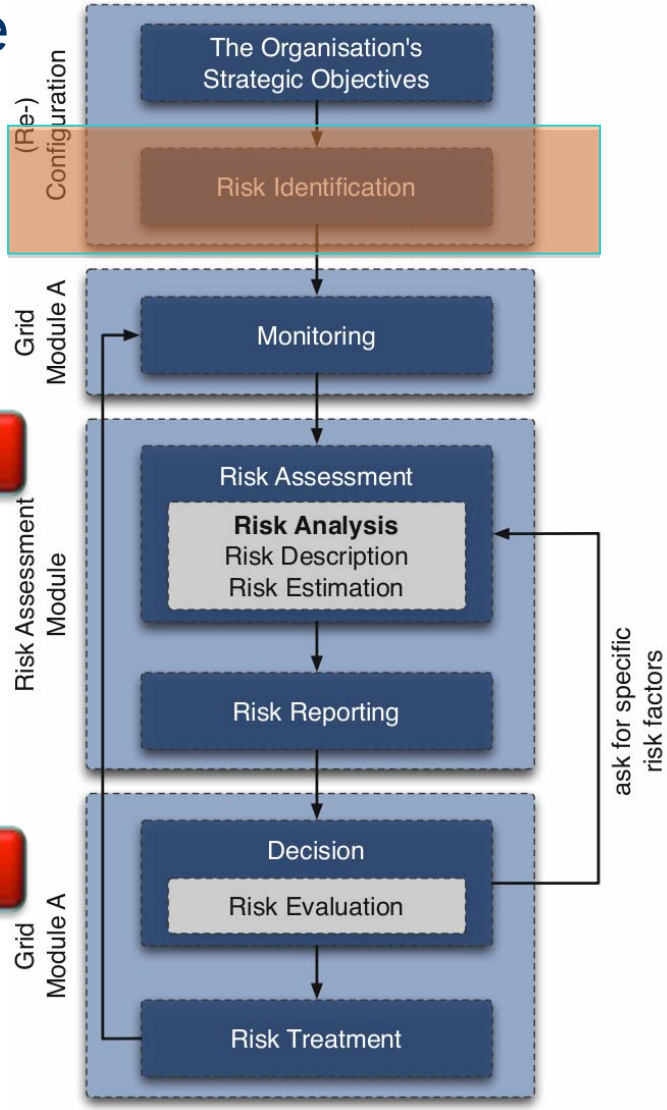
- Define strategic objectives
  - maximize profit
  - maximize reliability
- Define policies
  - minimum profit margin
  - maximum acceptable Probability of Failure (PoF)
  - maximum expense of fault tolerance (FT) mechanisms
  - if not all SLAs can be fulfilled which should be violated
    - those with lowest profit
    - keep those of good customers



Risk Identification

- Risk Identification: basis to reliable and accurate risk estimations
- Identify threats and their origin
  - SLA violation due

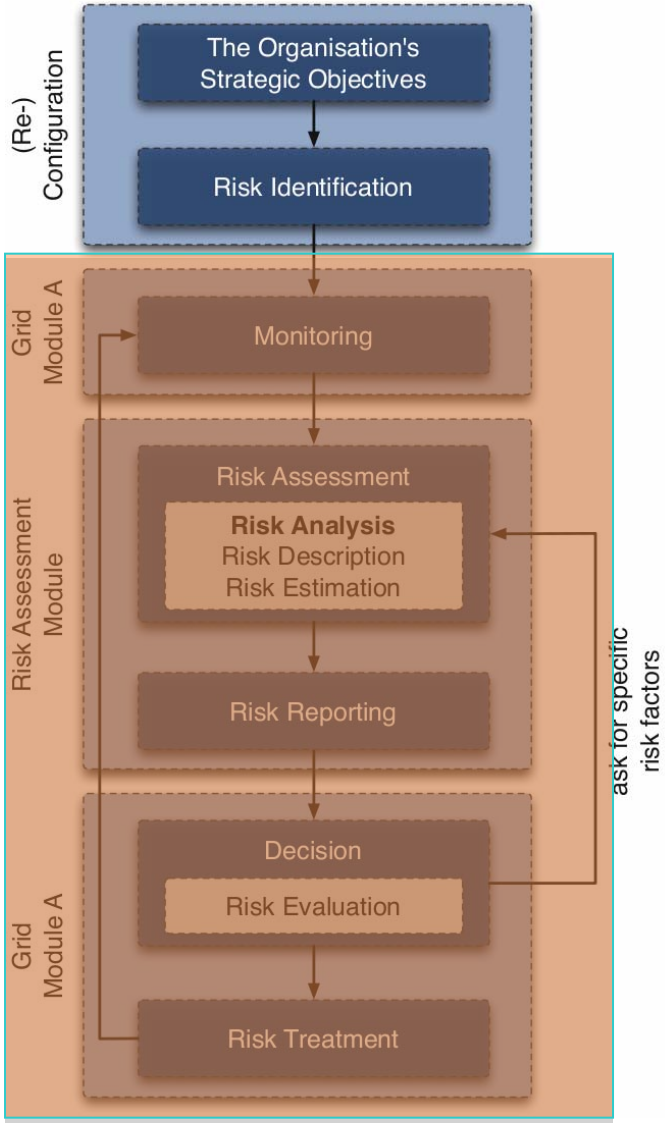




- Define responsibilities for failures

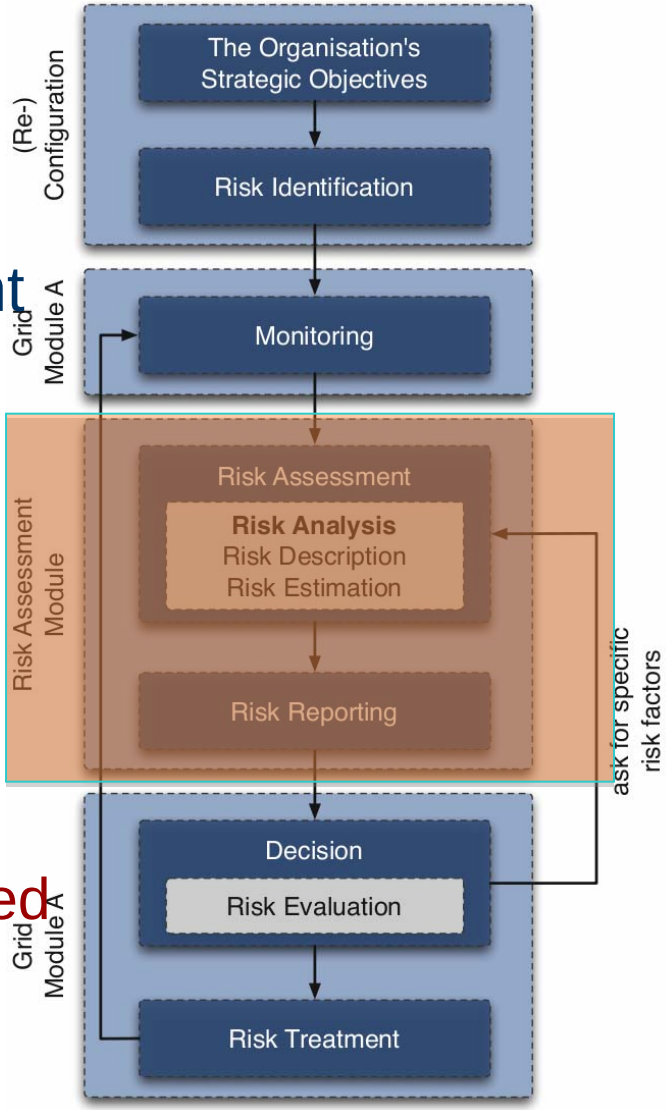
Grid Modules

- Responsible for decisions and risk treatment are several modules
  - Negotiation Manager
    - pre-selection of SLA offers
      - possible acceptable
      - direct reject
  - Scheduler
    - resource allocation
    - and reservation
  - Fault Tolerance Manager
    - planning FT mechanisms
    - react to PoF changes
  - Security Manager
    - probability for DOS attacks



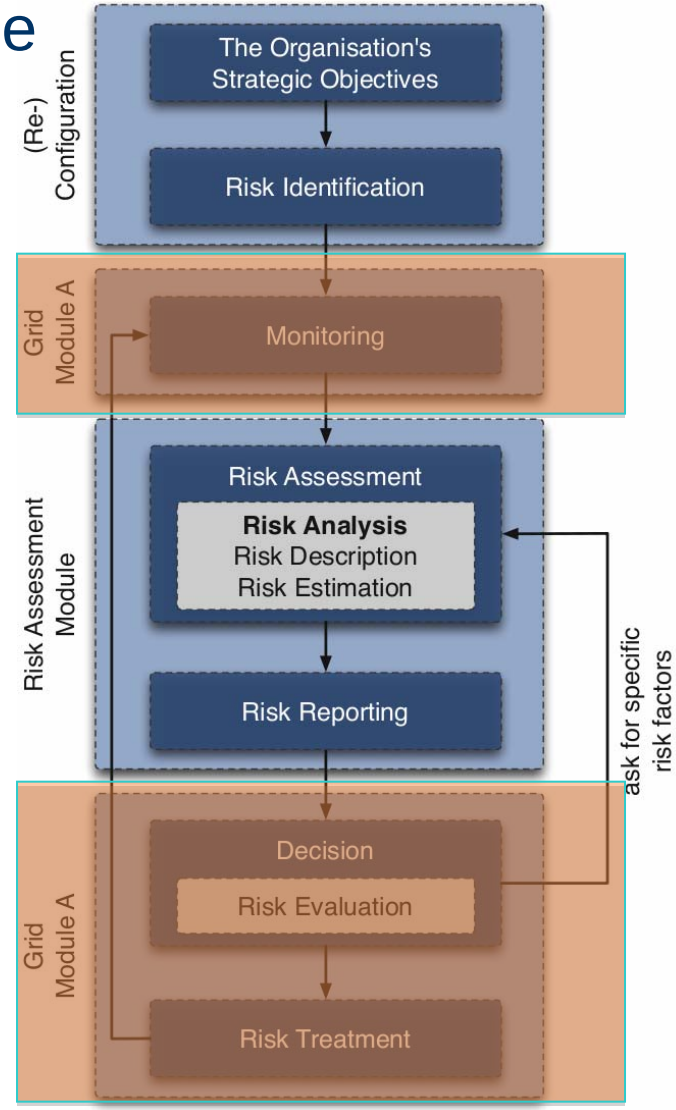
Input Specification of Risk Factors

- Minimal required input parameter
  - for initial risk
- For more accurate risk assessment
  - optional input parameters
- Targeted risk process
  - risk is queried by a specific module
  - reporting notifies this module
  - only authorized modules were informed about events



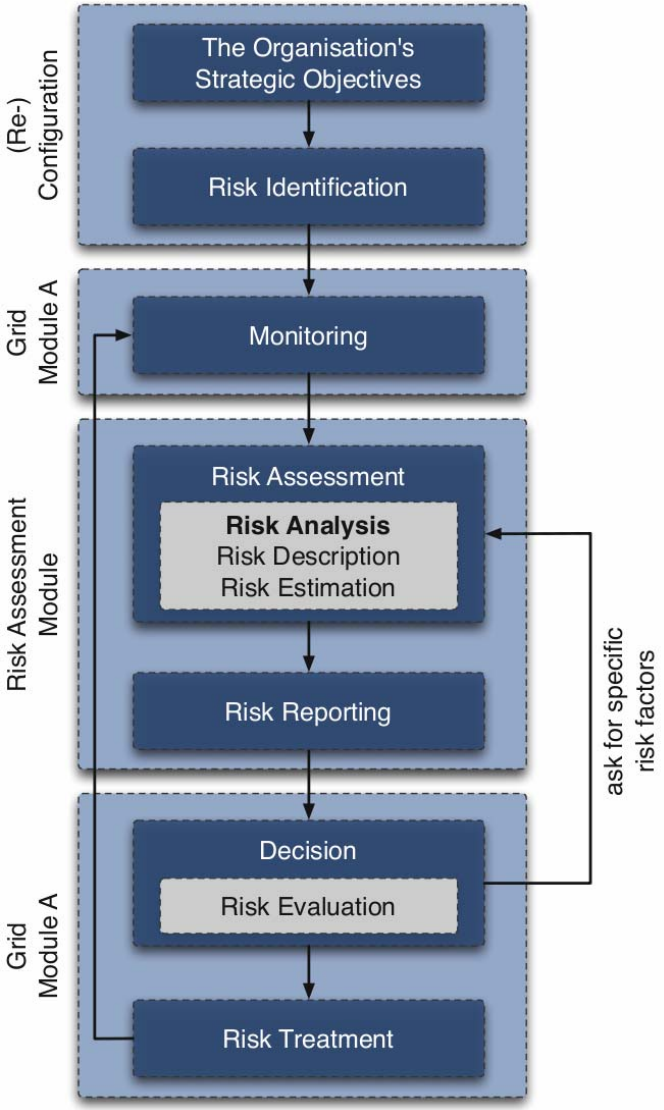
Decision Making and Risk Treatment

- Distinguish negotiation and runtime
- Decision making
  - accept SLA or not
  - select FT feature
- Failure management
  - performed after resource outage
    - initiate FT mechanisms
    - accept SLA violation
- Risk treatment
  - on single or all jobs
  - risk treatment might change PoF



Risk Review

- Will be performed periodically
  - several RAs in parallel
- Not coupled with the targeted RM process
- Compare estimated PoF with
  - actual occurrence of events
  - monitored in the Grid
- If residual risk is not as intended
  - adapt risk assessment



Conclusion

- Risk Assessment and Management is required for SLA provisioning
- Standard RM Processes cannot be used
- Grid RM Process has to run automatically
  ➔ little modifications necessary
- AssessGrid provides reference implementation which can be configured according to the organization's objectives and system

www.assessgrid.eu

References

- [Khali 06] Khalili, O., He, J., Olschanowsky, C., Snavely, A., Casanova, H.: Measuring the Performance and Reliability of Production Computational Grids. In: GRID 7th IEEE/ACM International Conference on Grid Computing (GRID 2006), Proceedings, Barcelona, Spain, IEEE (Sep 2006) 293-300
- [Dumi 05] Dumitrescu, C., Raicu, I., Foster, I.T.: Experiences in Running Workloads over Grid3. In: Grid and Cooperative Computing - GCC 2005. Volume 3795 of Lecture Notes in Computer Science., Beijing, China, Springer (2005) 274-286
- [Iosu 05] A. Iosup, M. Jan, O. O. Sonmez, and D. H. J. Epema. “On the dynamic resource availability in grids”. In: Proceedings of the 8th IEEE/ACM International Conference on Grid Computing (GRID 2007), pp. 26–33, Austin, Texas, USA, Sep 2007.


Thank you!