Data storage architecture to support various research needs and data lifecycle management

Ville Tenhunen, Minna Harjuniemi

13.6.2013, EUNIS Congress, Riga

Contents

- UH in nutshell

- Various needs

- Architecture and lifecycle management

- Some solutions

- Conclusions and discussion

UNIVERSITY OF HELSINKI

ESTABLISHED 1640

MULTIDISCIPLINARY INTERNATIONAL RESEARCH UNIVERSITY WITH STRONG SOCIETAL COMMITMENT

BILINGUAL (FINNISH AND SWEDISH), TUITION ALSO PROVIDED IN ENGLISH

MEMBER OF THE LEAGUE OF EUROPEAN RESEARCH UNIVERSITIES (LERU)

36,500 STUDENTS

5,800 DEGREES AWARDED ANNUALLY

8,500 STAFF, OF WHOM 4,800 RESEARCHERS/TEACHERS

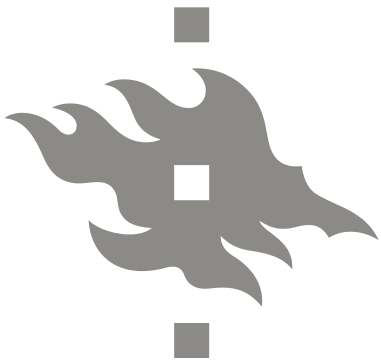**TOTAL ANNUAL BUDGET €650 MILLION**
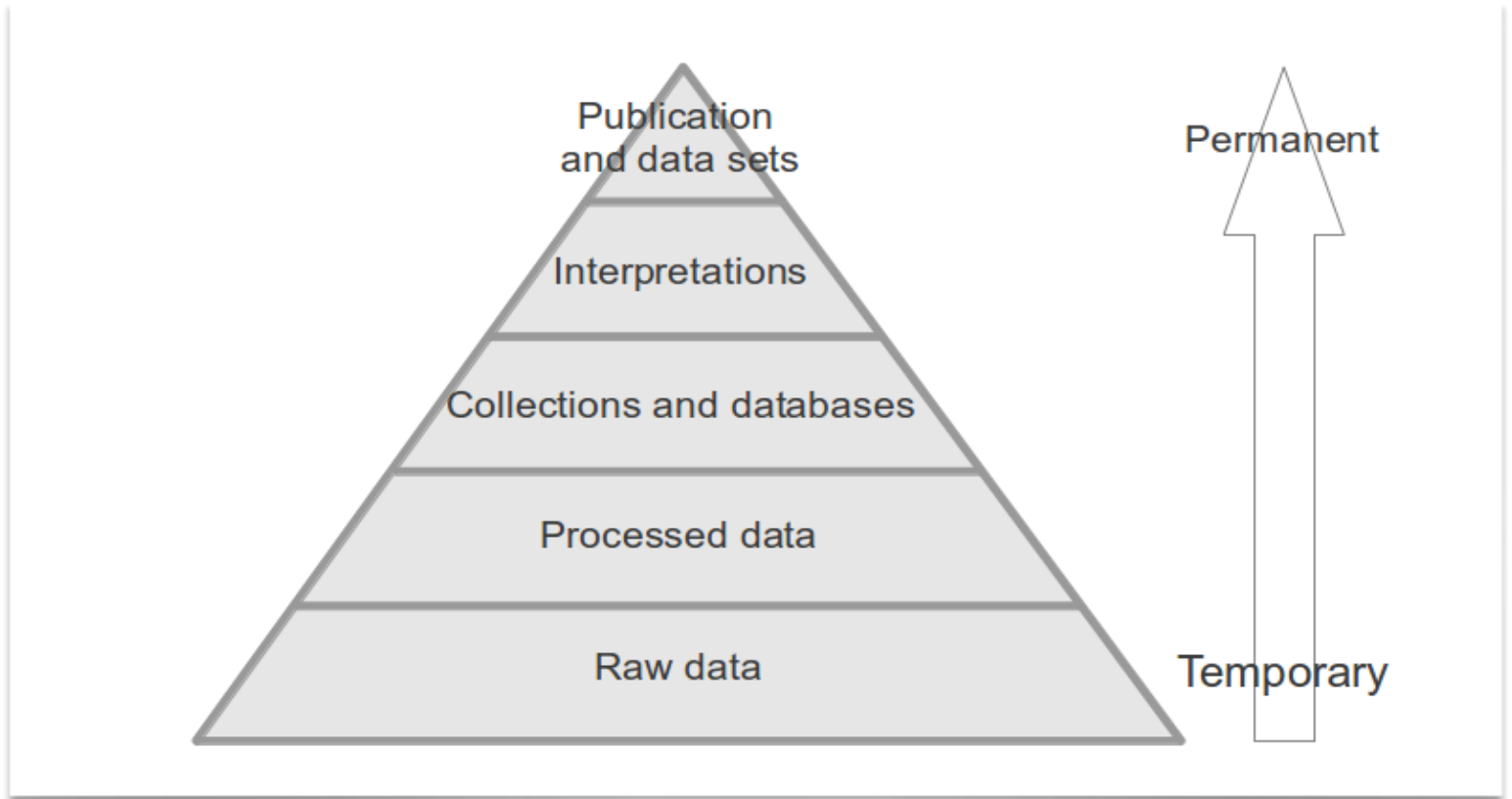
UNIVERSITY OF HELSINKI

Multiple changes in research data management

- Modern scientific infrastructures generate more data
- Researchers have applications and methods to use large amounts data
- New research areas have ability to use data intesive methods
  - They have their own traditions, methods and equipments
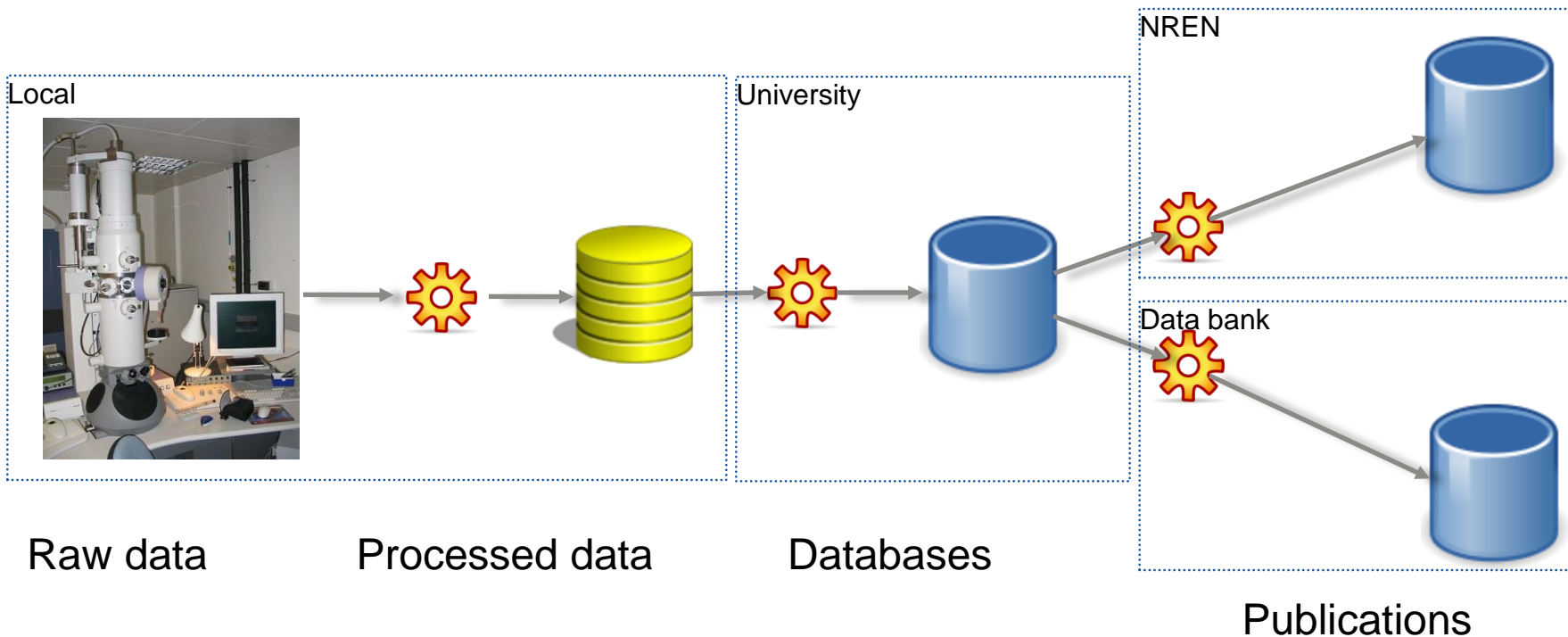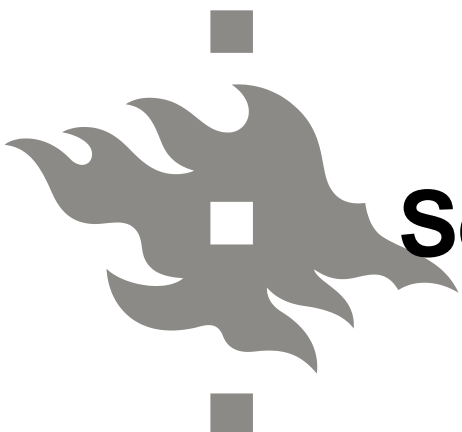 - Variation of skills and knowledge

Lifecycle of research data

Pyramid from bottom to top: Raw data, Processed data, Collections and databases, Interpretations, Publication and data sets. Arrow from Temporary (bottom) to Permanent (top).

Paths of data, example

NREN

Local

University

Data bank

Raw data   Processed data   Databases

Publications

Picture of the microscope: Wikimedia/David J Morgan from Cambridge, UK
www.helsinki.fi/yliopisto
6   6

Selection criteria of solution

- Amount of data?

    - Capacity need now? Next year? After two years?

- Lifecycle of data (when data should be deleted)?

    - Less than one year? 5 years? 20 years? More?

- How data would be used?

    - Through Windows desktop, Linux/Windows server application or some other way?

    - Does researcher process data in the storage (i.e. need of fast disks or not)

    - Storage for raw data, processed data, shared data or archive?

... and some more

- Do researcher want to share data?

  - Have researchers UH's access rights or possibilities to use federated logins (HAKA, Kalmar union, eduGAIN?

- Is data classified?

  - e.g. personal data, patents

  - Data security issues

- Do researcher need backups?

- Budget?

  - Investments, maintenance

  - Remember the lifecycle
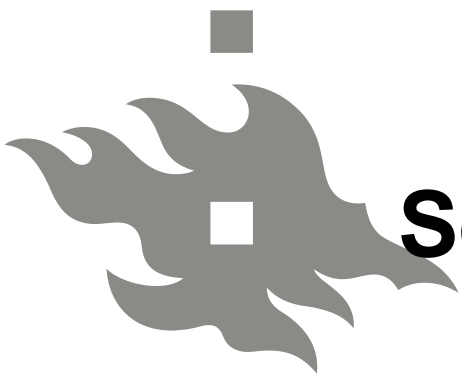
Solutions at UH

• Small storages(< 1 TB)

    –   Group storage (NAS)

    –   Web services

• Mid size storages (1-100 TB)

    –   IT Center's NAS and SAN

    –   Tailored solutions

    –   Local storage servers

• Large scale storages (> 100 TB)

    –   IT Center's NAS and SAN

    –   Tailored solutions (infrastructure, networks, planning etc.)

NAS = Network Access Storage, SAN = Storage Area Network

Local storage server - "Good enough"

- IT Center will arrange for a framework agreement with 3 vendors for purchasing storage servers

- Size ~ 10 - 120 TB

- Normal servers or efficient servers for computing

- Comments from users:"Good enough for many purposes"
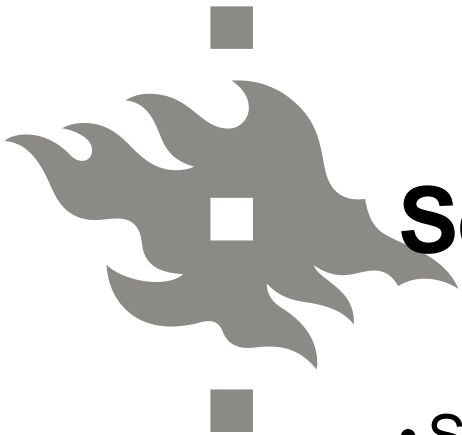
Services outside UH

- Services of CSC - IT Center for Science (owned by Ministry of Education and Culture)
    - Also VLAN between UH and CSC
    - Lightpath between networks
- Other research networks
- Cloud services (for example Amazon Glacier, Microsoft Azure)
    - Security issues
    - Administration like user accesses

NREN = National research and education network
VLAN = Virtual Local Area Network

Technology is not everything

Additionally it is very important to take care of for example these things:

- Meta data, usability and accessibility

- IPRs, publicity and other regulations

- Sponsors and ownerships of the data

- Research communities and the priciples of science

- Motivation of researchers is key aspect (and merits for researchers)

- Data security

Conclusions

- One size does not fits all

- "Good enough" is sometimes enough

- Budget is usually guiding factor

- Research data and it's management have to be enabler of the research, not a limitation

- The researcher knows exactly the needs she/he has, but not always knows what is the best solution

- Let's keep it simple
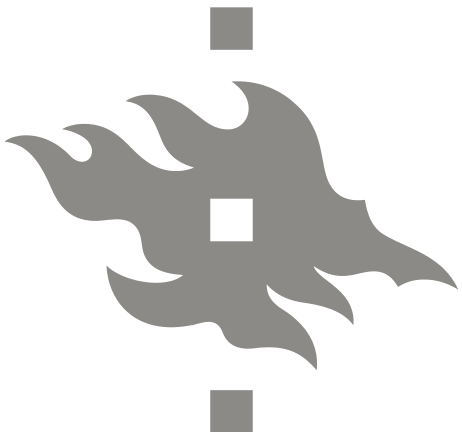
Thank you!

ville.tenhunen@helsinki.fi

minna.harjuniemi@helsinki.fi