Norsk Regnesentral
NORWEGIAN COMPUTING CENTER

www.nr.no

On Long-Term Archiving, Forensics, Security and Privacy Challenges

Habtamu Abie, Dr. Scient.

Senior Research Scientist, DART

Norwegian Computing Center

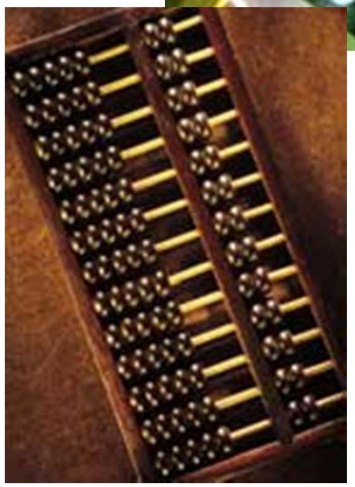
MultiTeam User Conference, 5 April, 2011
Soria Moria Conference Center, Oslo

Outline

- NR – Norsk Regnesentral/Norwegian Computing Center
- Long-term Archiving
- Digital Forensics
- Security and Privacy
- Research challenges
- Summary

Facts about NR

- Applied research (Private non-profit)
- Financed by
  - domestic private companies
  - public sector
  - Research Council of Norway
  - EU
  - international companies
- Established in 1952
- 65 research scientists
- Turnover 75 MNOK, 8,7 M EURO



EU projects
(6th and 7th Framework Programme) — examples

- GEMOM
  - Genetic message oriented secured middleware
- HATS
  - SW modeliing and IT-security
- DIADEM
  - Inclusive access for disabled or elderly persons
- CREDO
  - Modelling of evolutionary structures for distributed systems
- Geoland
  - Monitor land cover and vegetation
- CCII
  - Climate change and the insurance industry

ICT Research

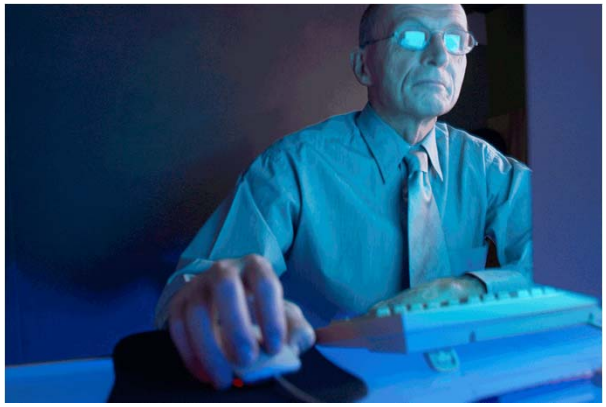
- Security
  - Privacy and Identity Management
  - Adaptive Security Measures
  - Risk Management
  - Trust Management
  - Security Analysis and Evaluation
  - Digital forensics
  - Digital Rights Management (DRM)

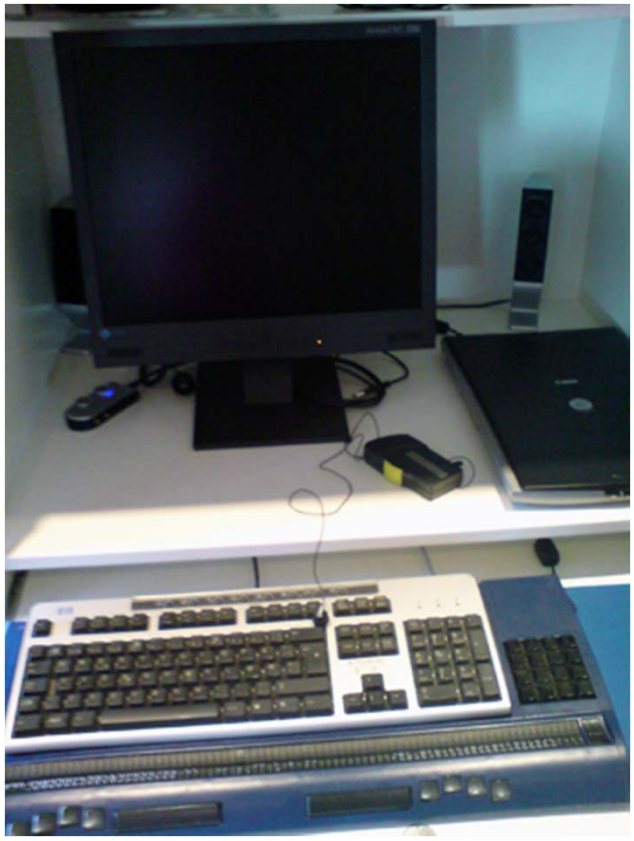
ICT Research ...

- Multimedia multichannel
  - Video/Audio Streaming
  - Multimedia Metadata & Databases
  - Mobility
  - Digital TV

ICT Research ...

- eInclusion
  - Universal design
  - Product and services accessible by as many users as possible





Work station for a blind person.

Background – An Introduction

- Long-term Archiving at NR
  - LongRec - Records Management over Decades (2007 - 2009)
  - MARIAGE - Making Rich Media Accessible for Generations (2007-2009)
- Digital Forensics at NR
  - Project on Digital Forensics in 2004 and 2005
  - Research exchange program in 2008
  - Paper in the Journal of Forensics Sciences in 2010 [Trcek 2010]
- Security and Privacy at NR
  - Several Security and Privacy related projects
  - PETWeb II -- Privacy-respecting Identity Management for e-Norge (June 2009 - May 2013)
  - PETweb - Privacy Enhancing Technologies for Web-based Services
  - PerProt - Personalized Internet-Based Services and Privacy Protection

Long-term archiving

- Main objective
  - Persistent, reliable and trustworthy long-term archival of digital documents, with emphasis on availability and use of archived documents

- Challenges
  - to establish theory, mechanisms, and technology that enable companies to trust long-term (several decades) storage of digital original documents, and
  - to be able to use and update the documents throughout their lifetime

How to make digital objects trustworthy?

- Object (document) to be archived
  - Content data object
  - Representation information
- Additional trust-enhancing information
  - Provenance information, documenting
    - creation of target object
    - any alternations in content/format over time
    - chain of custody
  - Data object identification
    - Unique identifiers
  - Authenticity and integrity-validating information
    - checksums, digital signatures, digital watermarks

How to make archival systems trustworthy?

- Access management
- System/repository security
- Repository redundancy
- Trustworthy work processes

- Use proactive risks analysis
  - predict future changes
  - measure regularly and adaptively the level of trust and security that has been achieved, and can be achieved

When to include trust-enhancing meta data

- When digital objects are entering an archival system
  - Specifically information from the time of its creation to the time it enters the repository
- Inside the archival system, whenever
  - Anyone modifies the digital object or the meta information
  - Format conversion might be necessary over time
- Encapsulate in XML-based schemas
  - Add, do not delete information
  - Used digital signatures, checksums, etc.

Two trust strategies

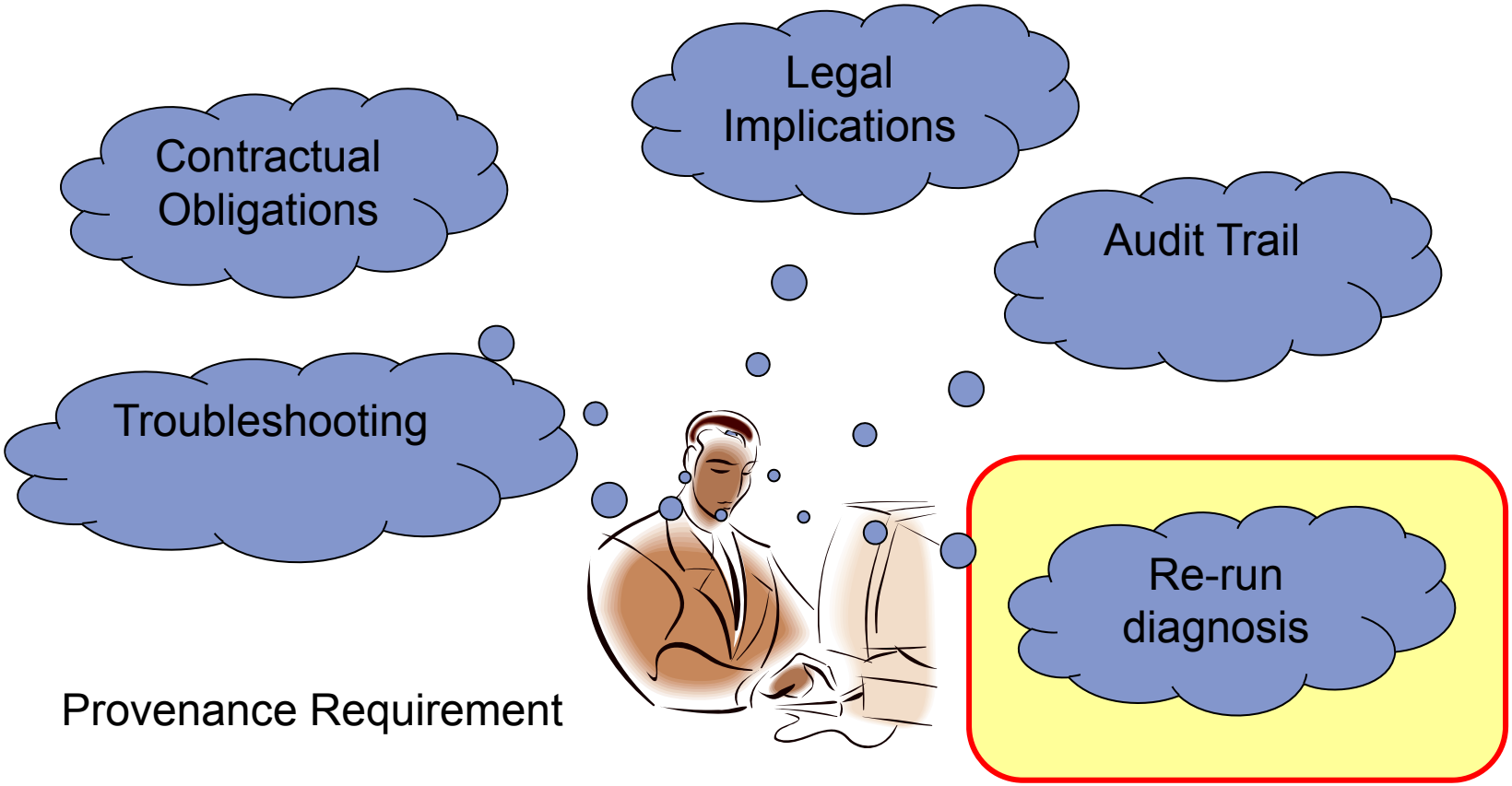
- Optimistic
  - Assuming evidence will not be needed
  - Do not put too much additional effort into increasing trustworthiness
  - Take the risk of not changing data formats, adding metadata, etc.
  - Less expensive choice, if evidence is not needed in the future

- Pessimistic
  - Assuming evidence is needed in the future
  - Add successively metadata/evidence in every life-cycle phase
  - Use cryptographic seals whenever suitable
  - Additional cost in daily business is a consequence
  - But we are prepared if anything happens (!)

Concept and utilization of provenance

- Provenance of data is the process that led to that piece of data
  - represented by some suitable documentation of the process (i.e. workflow execution) that led to the data.
- Distinguish between a specific piece of information documenting some step of a process from the whole documentation of the process
  - the former is referred to as a p-assertion, which is essentially an assertion made by an actor pertaining to any aspect of a process
  - the documentation of a process would therefore consist of a set of p-assertions made by all the actors involved in that process
- Archival facility and re-signing/re-encrypting provenance information periodically over the long-period storage

Provenance data users



Source: DAME (Distributed Aircraft Maintenance Environment)
http://www.cs.york.ac.uk/dame/OpenDayProvPresentationV0.3.ppt

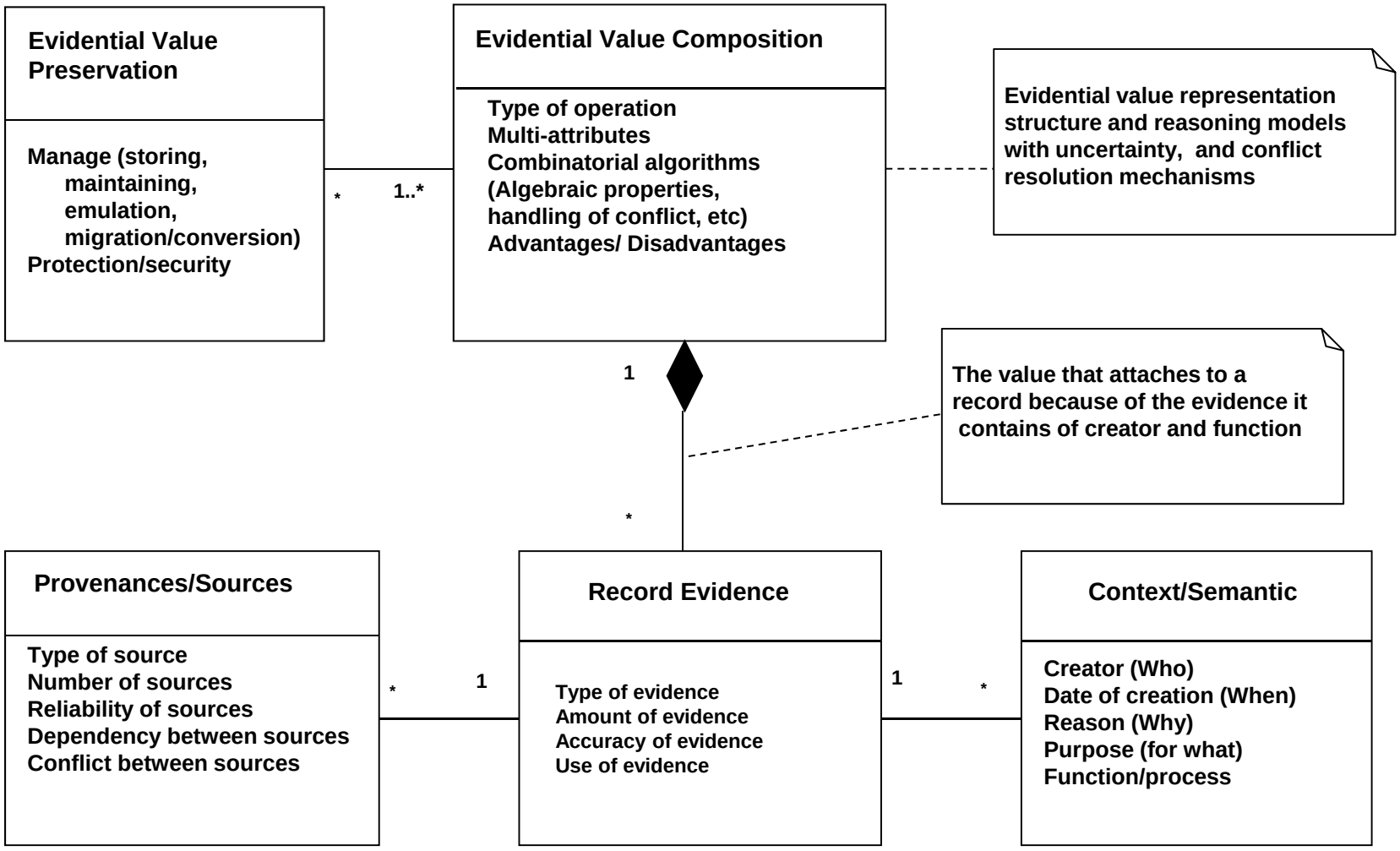
Preservation of trust and security

- Confidence in preservation of
  - availability (to authorized actors)
  - integrity (correctness)
  - confidentiality (to unauthorized actors)
  - protection of IPR (of ownership)
  - accountability (traceability of actions and events related to the document)
- Use of evidential value of a record as an index for the degree of trust
- Maintaining persistent security services over time

PhD in Assessment of the trustworthiness of digital records over time

- Establishing trust and security
  - use of evidential value of a record as an index for the degree of trust
- Evidential value
  - the quality of records that provides information about the origins, functions, and activities of their creator, relates the process of creation
- Validation of evidential value
  - degree of trust in document correctness: expressing, measuring, verifying and preserving
  - including document content, context, semantics, presentation, and trust management
- Assessment of the trustworthiness of digital records based on their evidential values
  - applying a rigorous formal approach, the trustworthiness of digital records can be assessed objectively
  - Addressing temporal, conflict, weighting and dependency aspects

Preservation of evidential value



Forensics – current landscape

- SoA and the Internet of Things constitute changes that imply
  - dynamic IT services, hence
  - dynamic security services

- Which again requires dynamic (and complex)
  - Forensics,
  - Security, and
  - Privacy **mechanisms**

in order to produce admissible evidence (in court)

Forensics – aspects

- Forensics process involves
  - Collection
  - Preservation
  - Analysis
  - Presentation
- Forensics Readiness
  - about what information is retained for forensics purposes as potential digital evidence, proactively
- This must cover several aspects
  - Global, national, enterprise, individual, technology

Note: It is also by nature a multidisciplinary science !

Forensics – activities

- Admissible evidence is created by forensics with two related activities
  - Digital Forensics Readiness (DFR)
  - Forensics Investigation
- DFR
  - everything we do up to the point where we start to investigate
  - many investigations do not lead to a court case where evidence has to be presented
  - investigation: post mortem analysis of digital traces by increasing the availability and quality of the raw traces
- Both activities depend on each other
  - Here we focus on DFR

Forensics – DFR challenges

- Traditional crime => investigate => physical evidence
- Cyber crime => investigate => no physical evidence
- Traditional IT investigations
  - Fixed infrastructure, fixed SW
  - Large logs, in physically protected stores, logical protection with MACs and stored almost everything
  - Extract almost “everything” by secure procedures, then analyze the event and maintain the “chain of custody”
- SoA and IoT investigations
  - Changing infrastructure, dynamic services, limited memory/storage
  - Small logs, copy to unprotected stores during run-time, can potentially be aggregated in a different location
  - Extracts are more fragmented (so it must be the important information) and the chain of custody is much more fragile

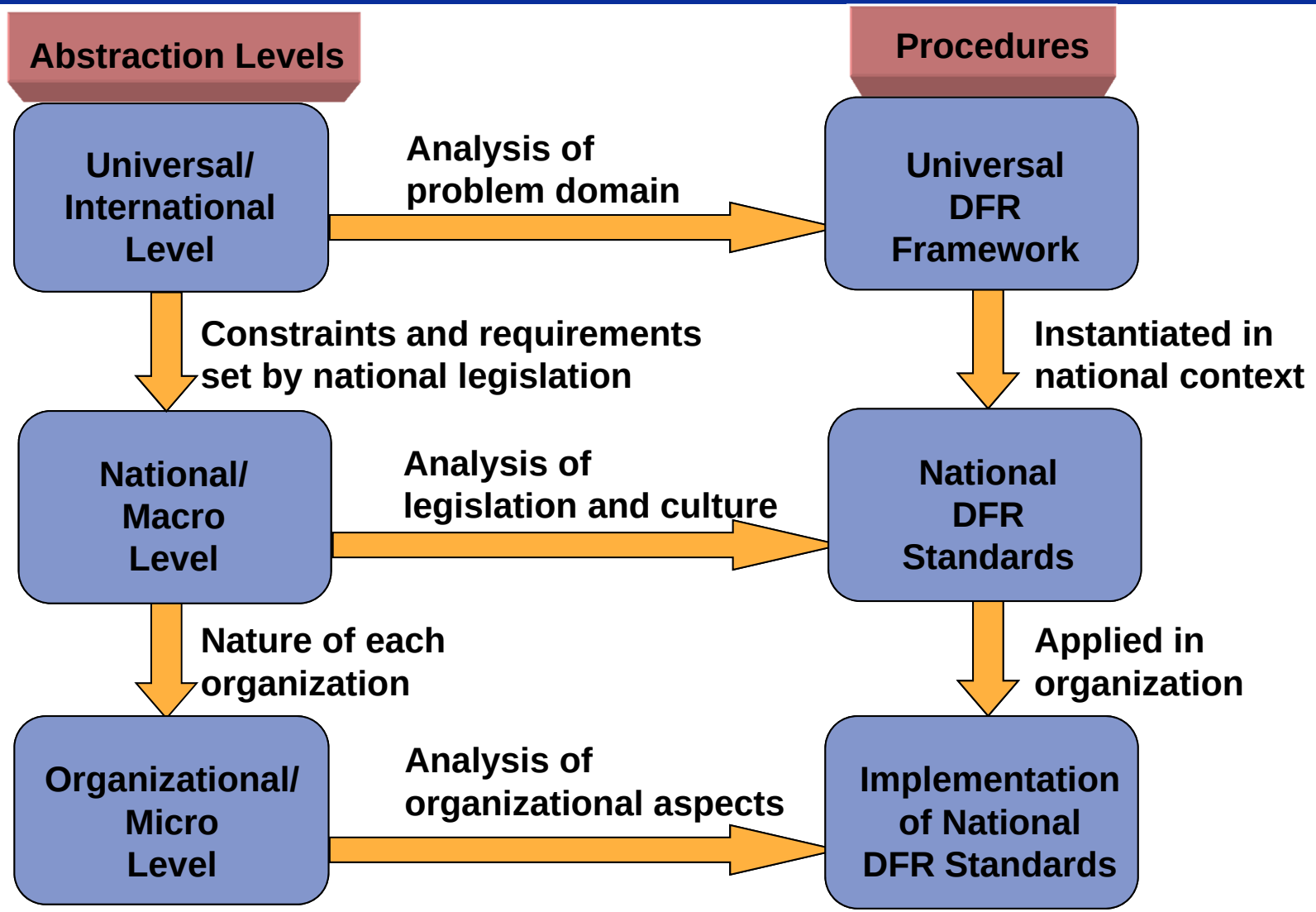
Forensics – DFR Framework goals

- Facilitate good practices and technology that will
  - identify the responsible
  - through a chain of admissible evidence
  - without infringing on individual rights
- In a “top-down” approach, where the goal is to have a
  - **Template legislation**
- **Template legislation** that is
  - Internationally recognized
  - serves as a “model law”
- Foster harmonised national implementations

Forensics – Framework principles

- Initial DFR approaches focused on
  - Logging techniques, IDS data usage, acquisition, evidence handling…
- **Rowlingson** [Rowlingson 2004] published the now well known paper "***A ten step process for forensic readiness" -*** Active collection of potential evidence
- All this complexity, and in addition one should cover
  - Holistic aspects of forensics
  - Several levels of detail and work processes
  - DFR (Technical) policies for system configuration
- Misc. aspects covered in the literature are the relationships to
  - existing response plans, sound investigations
- But still no agreed upon approach to DFR within organisations

Forensics – DFR Framework

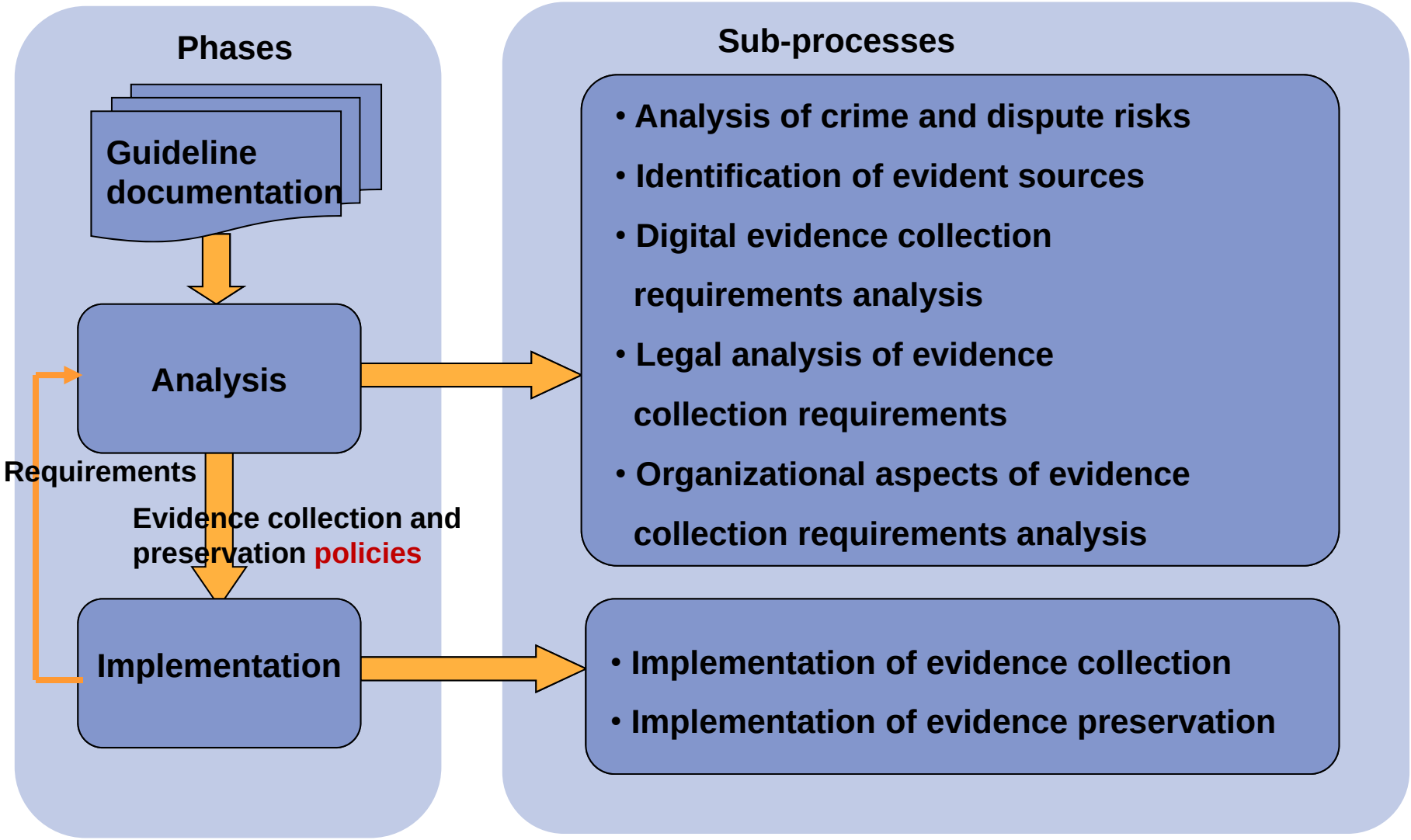


Forensics – process model

- Universal level
  - guidelines for preservation, without violating privacy, identification and classification of evidence sources, guidelines and standards for reporting incidents.
  - e.g., by the minimal set of requirements on privacy
- National level
  - analysis of restrictions from national legislation, potential constraints on DFR procedures, what is lawful collection
  - e.g., by concrete legislation on privacy issues
- Organisational level
  - analysing the organisations need, capabilities and exposure
  - e.g., by concrete procedural coverage of privacy issues between employer and employees

Then we add the “process view”

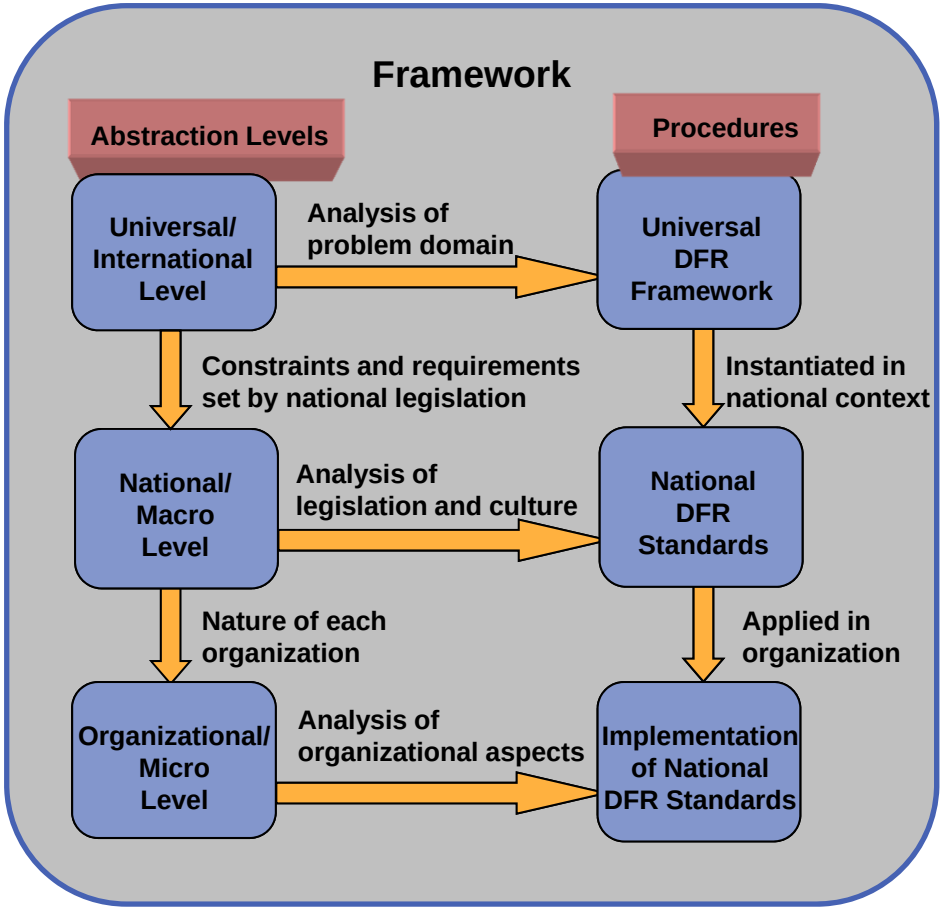
Forensics – process model / view



Forensics – from concept to operations

- Need to translate the overall policies to
  - actual security **services** (e.g. Integrity)
  - each implemented by a certain security **mechanism** (e.g. Integrity by SHA-256 with Qualified Certificates)
- Here we stop at the abstraction level
  - where the main property of a certain implemented mechanism is its “strength”.
- This makes the framework
  - more flexible and able to adapt to changing “confidence” in different cryptographic algorithms and the like …

Forensics – Framework mapping



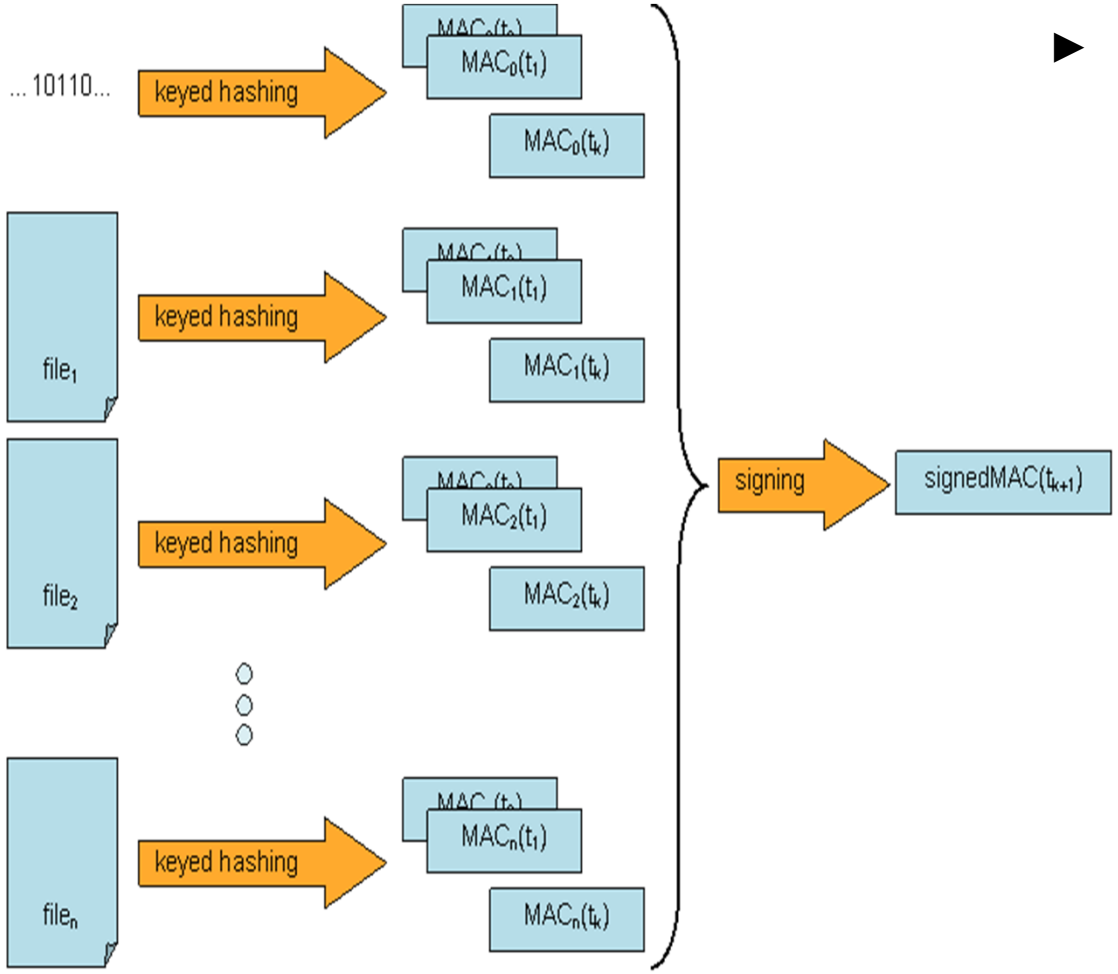
Policies

Forensic Security Services

Mapping

Forensic Security Mechanisms

Leveled MACs for integrity assurance



- Level Message Authentication Codes (MACs) -
  - to assure integrity of the original forensic data and its copies
  - the first k time-intervals keyed MACs are produced, which are afterwards digitally signed every (k+1)-th interval
  - digital signatures are applied only at the second level (or at the third level) on the MAC that is derived form the MACs generated at lower level

Threats to privacy in the forensic analysis [Stahlberg 2007]

- Evaluation of several real database systems
  - reveals that deleted data is not securely removed from database storage and that users have little control over the persistence of deleted data
- The problem of unintended data retention must be addressed by proposing a set of system transparency criteria
  - data retention should be avoided when possible, evident to users when it cannot be avoided, and bounded in time
- Specific techniques for secure record deletion and log expunction
  - must be developed that increase the transparency of database systems, making them more resistant to forensic analysis

Privacy protection guidelines

Ten guidelines [Srinivasan]

1. Remove personally identifiable data from storage media
2. Store an identical copy of any evidentiary media given to law enforcement
3. Limit search to goal of investigation
4. Handle time stamped events in strictest confidence
5. On networks, packet acknowledgement be via the use of tokens than IP addresses
6. Safe storage of all internal logs
7. Preservation of event logs in external nodes
8. Put policies in place for actionable items related to attacks
9. Put policies in place for safeguarding backed up data related to an investigation
10. Handle disposal of sensitive data in a secure manner

Research challenges

- How to instantiate the theory from the Framework
  - Collect and integrate best-practice guidelines
  - Demonstrate in a IoT environment
  - Demonstrate it for evidence handling a court
- What Privacy means in the future (IoT)
  - Vast number of identifiers requires control
  - Partially uncontrolled environments creates risks
  - What can, or should, be Anonymous
  - What influences our (legal) notion of “privacy”
- Harmonization of forensics, security and privacy enhancing technologies
  - So that all stakeholders (except abusers) are winners

Long-term archiving summary

- Evidence
  - is additional information about how a digital object/document is created and kept
  - can be collected successively
  - can be cryptographically sealed
  - can be represented internally (XML) together with the document
  - can be stored in more than one place
- need to establish trust between private collections and repositories
- Possible to assess the trustworthiness of digital records objectively using formal approach
  - Using the records' evidential values as a measure of trustworthiness

Summary…

- Digital Forensics
  - A Framework for DFR
  - Cover new IT environments
  - Procedures; legal -> organisation -> technical
  - Mapping to Security services
  - Structure with flexible implementations
  - Policies should be in place to protect privacy of subjects in an investigation
  - Computer Forensics provides plenty of trace back capabilities
- Security and privacy issues
  - Some are integrated, but not all
  - Many interesting research challenges ahead

References

- LongRec Project http://www.longrec.com/Pages/Default.aspx
- [Rowlingson 2004] Rowlingson R. A ten step process for forensic readiness. Int J Digit Evid 2004;2(3):1–28.
- [Trcek 2010] Denis Trcek, Habtamu Abie, Åsmund Skomedal and Iztok Starc, Advanced Framework for Digital Forensics Technologies and Procedures, Journal of Forensic Sciences, Volume 55, Issue 6, pages 1471–1480, November 2010
- [Stahlberg 2007] Patrick Stahlberg, Gerome Miklau, and Brian Neil Levine, Threats to Privacy in the Forensic Analysis of Database Systems, SIGMOD'07, June 11–14, 2007
- [Srinivasan] S. Srinivasan, Security and Privacy in Computer Forensics Applications, http://www.utm.edu/staff/jclark/midsouth/Srini.ppt

Contact & Visit

Thank You for Your Attention!!

With many thanks to Arne-Kristian Groven (groven@nr.no) and Åsmund Skomedal (skomedal@nr.no) at Norsk Regnesentral for their contributions

E-mail: habtamu.abie@nr.no

Visit: www.nr.no/dart

The End!