Digital Curation as a Key Component in Research Infrastructures:
From Data Preservation to Processes Preservation and Verification

Andreas Rauber

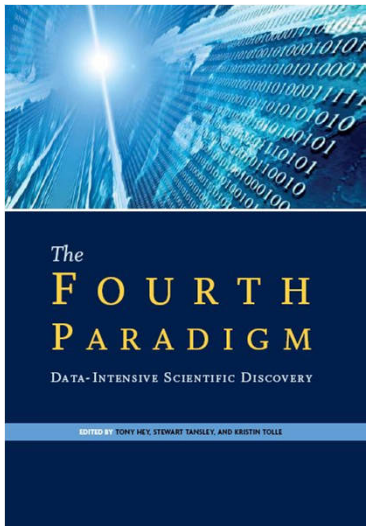
Department of Software Technology and Interactive Systems

Vienna University of Technology
rauber@ifs.tuwien.ac.at
http://www.ifs.tuwien.ac.at/~andi

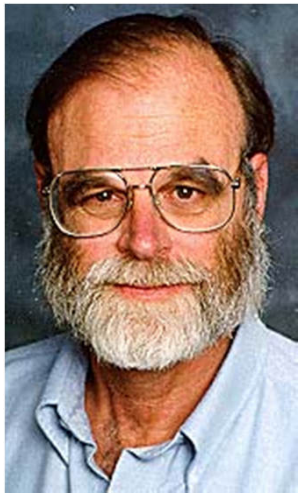
Outline

- Data-driven Science
- Digital Preservation
- From Data to Processes
- Technical solution: Process Capture
- Contractual solution: Holistic ESCROW
- How to evaluate?
- Data Curation and the Role of Digital Libraries

The Fourth Paradigm



- The Fourth Paradigm
  Tony Hey, Stewart Tansley, and Kristin Tolle (Eds.),
  Oct. 2009, Microsoft Research
  http://research.microsoft.com/en-us/collaboration/fourthparadigm/

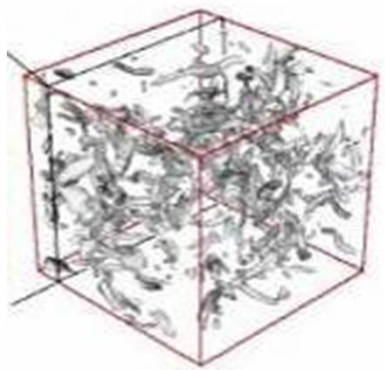


- Jim Grey (1944-2007)
  Turing Award Winner 1998
  - Presentation on Jan. 11.2007, NRC-CSTB
    http://research.microsoft.com/en-us/um/people/gray/talks/NRC-CSTB_eScience.ppt
  - Transcript:
    http://research.microsoft.com/en-us/collaboration/fourthparadigm/4th_paradigm_book_jim_gray_transcript.pdf

The Fourth Paradigm

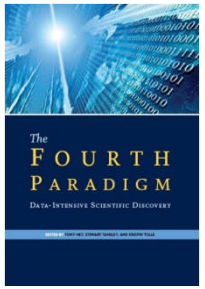
1) Empirical Science
   - ~1.000 years ago
   - Description of observed phenomena
2) Theoretical Science
   - ~100 years ago
   - Model buidling, Generalization
3) Computational Science
   - ~10 years ago
   - Simulation of complex Phenomena

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{4\pi G\rho}{3} - \mathrm{K}\frac{c^2}{a^2}$$



(adopted from Jim Gray, eScience Talk at NRC-CSTB meeting
Mountain View CA, January 11, 2007, slide 4)

The Fourth Paradigm



4) Data-intensive Science

- ~ now
- Connects theory, experiment and simulation
- Huge amounts of data from sensors and simulations
- Data processing via software
- Storing data in networked infrastructures
- Gaining knowledge by analysis of integrated data

- eScience, Data-driven Science, Data-intensive Science
- Studies / Meta-Studies, Integration
- Data is the key enabler --> need to preserve data

From eScience to DP

- Preserving data: where is the problem?
  - collecting the data
  - describing the data
  - curating the data: digital preservation
- What's the challenge in Digital Preservation?
  - storage and back-up?
  - anything else?
- The 3 levels of DP challenges:
  - physical, logical and semantic digital preservation

Outline

- Data-driven Science
- Digital Preservation
- From Data to Processes
- Technical solution: Process Capture
- Contractual solution: Holistic ESCROW
- How to evaluate?
- Data Curation and the Role of Digital Libraries

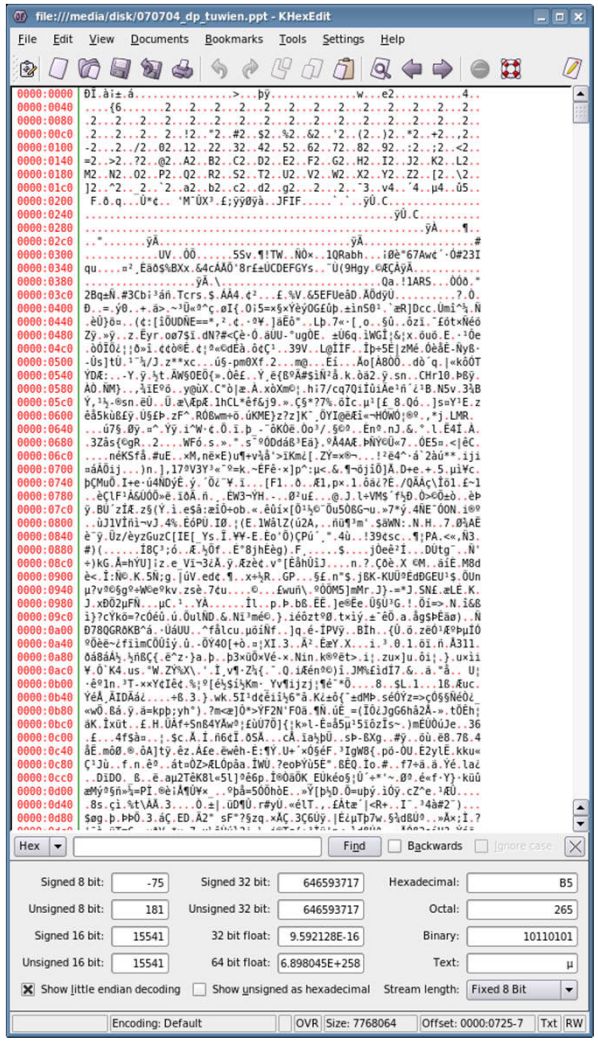
Why do we need Digital Preservation?

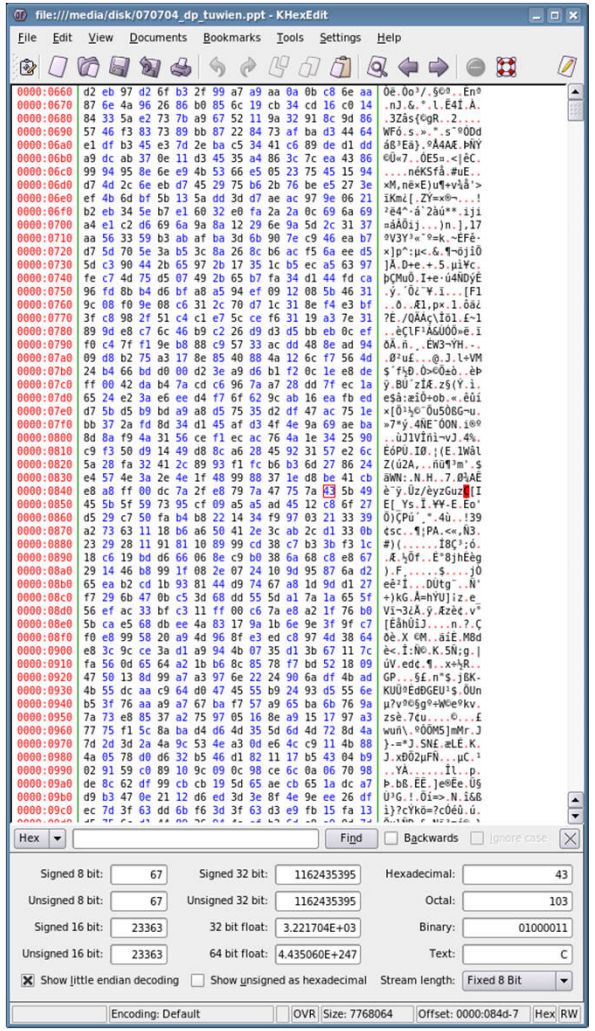
Why do we need Digital Preservation?

1. Physical Preservation (Bit-stream preservation)

- Transfering to current storage systems
  - note: transfer may not be trivial
    (file systems, encodings, relative references, copy protection,...)
- Ensure redundancy
  - technologically
  - geographic spread
- Access, security
- Error detection, recovery, disaster planning

Why do we need Digital Preservation?



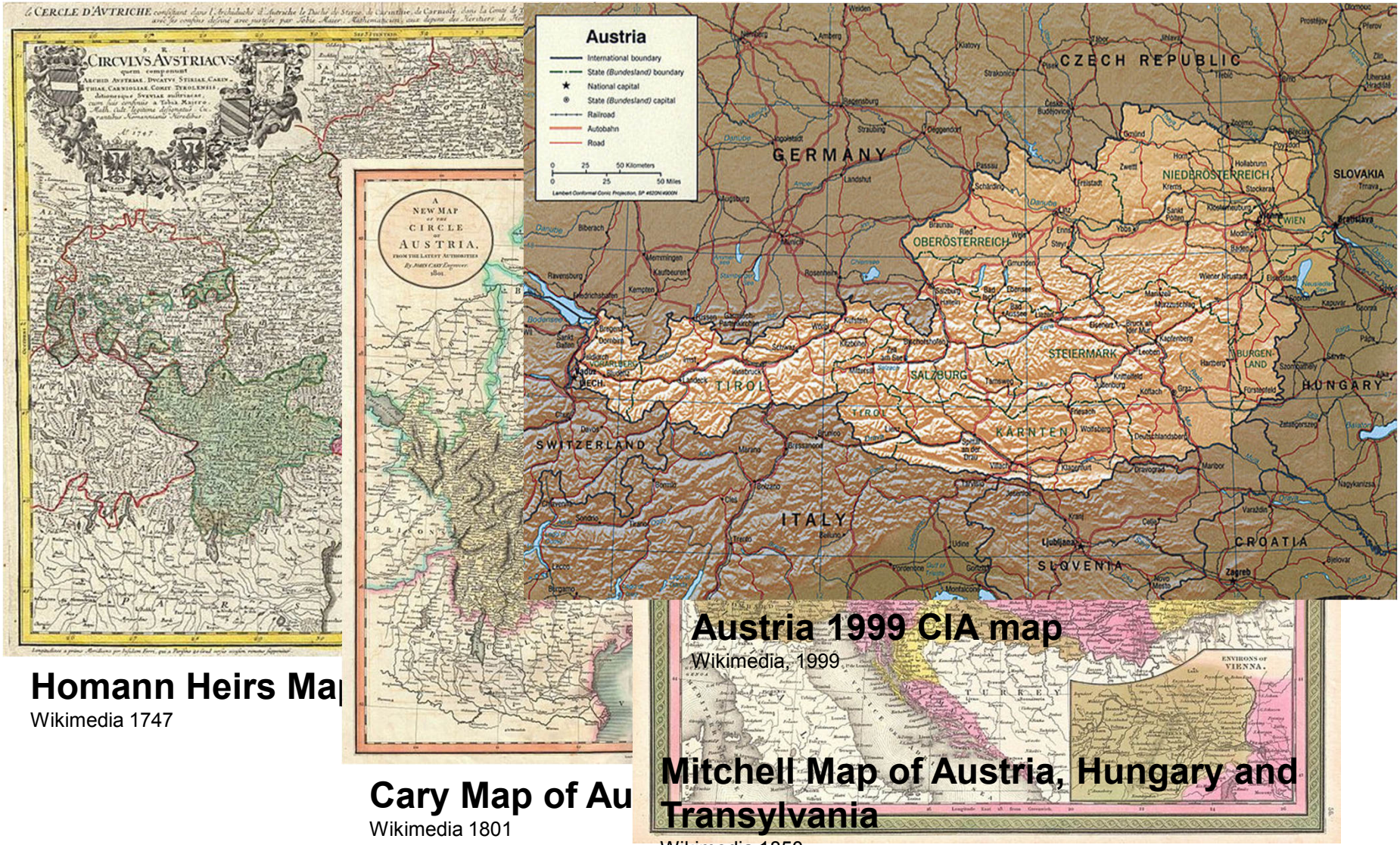


Why do we need Digital Preservation?

2. Logical Preservation

- Digital Objects require specific environment to be accessible :
  - Files need specific programs
  - Programs need specific operating systems (-versions)
  - Operating systems need specific hardware components
- SW/HW environment is not stable:
  - Files cannot be opened anymore
  - Embedded objects are no longer accessible/linked
  - Programs won‘t run
  - Information in digital form is lost
    (usually total loss, no degradation)
- Digital Preservation aims at maintaining digital objects authentically usable and accessible for long time periods.

Why do we need Digital Preservation?



Homann Heirs Ma
Wikimedia 1747

Cary Map of Au
Wikimedia 1801

Austria 1999 CIA map
Wikimedia, 1999

Mitchell Map of Austria, Hungary and Transylvania
Wikimedia 1850

Why do we need Digital Preservation?

3. Semantic Layer: information object

- How to interpret the data (information?) in the objects?
  - terminology changes:
    changes in country names, borders, connotation of words,...
  - concept changes:
    drunk driving: before 1998: 0.8‰ , afterwards 0.5‰
  - transformations: currencies/exchange rates, sensor resolutions,
  - provenance: actions applied to objects
    sources: who? / which sensor?, transformations, post-processing
  - context of objects:
    understanding the context of decisions, side-effects, quotations, calibration timestamps
- For preserving digital information, all 3 layers need to be addressed

Why do we need Digital Preservation?

- Essential for all digital objects
  - Office documents, accounting, photos, videos, emails, …
  - Scientific datasets, sensor data, databases, …
  - Applications, simulations, business processes, …

- All application domains
  - Cultural heritage data
  - eGovernment, Public administration
  - Science / Research
  - Construction industry
  - Health, pharmaceutical industry
  - Aviation, control systems, …
  - Private data
  - …

Why do we need Digital Preservation

3 levels of threat / preservation

1. Bit rot – physical preservation / bit preservation
   Physical Layer: how to keep the 0's and 1's
2. Object formats – logical preservation
   Logical Layer: how to remain able to open a file, run a program
3. Authenticity, interpretability – semantic preservation
   Semantic Layer: how to ensure we can understand/interpret data correctly

- What can we do?

Bit-level preservation

- Maintain bit-sequence
- Redundant storage:
  - Lockss: lots of copies keeps stuff safe
  - Cloud
- Distributed storage – physically separated
- Different technologies / platforms / production batches
- Controlled storage conditions
- Regular maintenance: type rewinding, disc spinning, …
- Maintain devices for accessing storage!
- Disaster recovery planning, monitoring and error detection
- Trade-off capacity, energy, effort

Bit-level preservation

Questions / discussion:

- How long do tapes / CDs / DVDs / HDDs / SSD last?
- What would be the most durable storage technologies?
- Is a DVD that lasts for 200 years a solution?
- What are the costs of bit-level preservation?
- What are the logistic challenges?
- Distribution and Trust?
- Are we allowed to store redundantly?
  - Copyright
  - Copy protection
- Am I allowed to store data in the cloud?

Why do we need Digital Preservation

3 levels of threat / preservation

1. Bit rot – physical preservation / bit preservation
   Physical Layer: how to keep the 0's and 1's
2. Object formats – logical preservation
   Logical Layer: how to remain able to open a file, run a program
3. Authenticity, interpretability – semantic preservation
   Semantic Layer: how to ensure we can understand/interpret data correctly

- How? What can we do?

Logical Preservation

Deja vue:

- Digital Objects require specific environment to be accessible :
  - Files need specific programs
  - Programs need specific operating systems (-versions)
  - Operating systems need specific hardware components
- SW/HW environment is not stable:
  - Files cannot be opened anymore
  - Embedded objects are no longer accessible/linked
  - Programs won‘t run
  - Information in digital form is lost
    (usually total loss, no degradation)
- Digital Preservation aims at maintaining digital objects authentically usable and accessible for long time periods.

Strategies for Logical Preservation

Technology Museum

- Keep the hardware (drives, computer,…)
- \+ Maintains full functionality
- \+ Creates time buffer to develop more permanent strategies
- \+ Requires detailed documentation of HW and SW, but this also helps
- \+ Only viable strategy for some types of objects? (which?)
- \- Economically and technically infeasible to maintain spare parts forever
- \- Requires huge "museum"
- \- Requires highly specialized know-how for all platforms and software

Strategies for Logical Preservation

Migration

- Transform into different format
- Continually or on demand (Viewer)
- \+ Widely used
- \+ Possibility to compare at time of migration
- \+ Resulting objects are always accessible
- \- Possibly undesired changes during migration
- \- Needs to be repeated again and again

Strategies for Logical Preservation

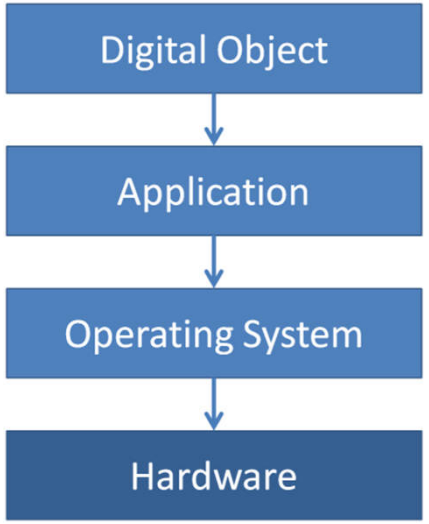
Emulation

- Emulation of Hardware or Software (OS, application)
- \+ Widely used principle
- \+ Many emulators available
- \+ Potentially preserving complete functionality
- \+ *Digital object is unchanged*
- \- *Digital object is unchanged*
- \- Complex technology, lot of research required
- \- Requires detailed documentation of the system
- \- Requires experience how to interact with emulated historic system in the future
- \- Emulators must be migrated as well
- \- Emulators potentially erroneous (Complexity)

Strategies for Logical Preservation

Excursion: Emulation vs. Migration

- Different on the pragmatic level, but conceptually identical
- Change occurs somewhere in the viewpath
- Have basically the same advantages/disadvantages and characteristics
- None of them guarantees identical rendering/performance of digital objects
- Many variants (e.g. viewer, virtualization)
- Need to be evaluated the same way



Strategies for Logical Preservation

Standardization

- Using open or de-facto standards
- \+ Simplifies DP process
- \+ Many tools available
- \+ Tools for standards are easier to build also in the future
- \- Significant effort required for standardization
- \- Loss at converting into standard
  (who is responsible?)
- \- Some object types cannot be standardized

Strategies for Logical Preservation

Standardization - Excursion into file formats
Proprietary vs. Open

- Proprietary
  - Documentation mostly not available
  - License and patent rules
  - License agreements subject to change
  - Restrictions for use and modifications may apply
- Open
  - Documentation available!
  - Unlimited use
  - No license fee
  - Open for modifications
  - No patent owners
- But: sometimes proprietary may better than open - **why**?
- Is the concept of "file formats" still useful?

Strategies for Logical Preservation

Data/Information Extraction

- Create abstract representation of information (e.g. databases or documents -> XML)
- \+ Independent of specific infrastructure
- \+ Many tools available
- \+ Easier to develop tools in the future
- \- High effort to develop tools for specific abstraction scenario
- \- Limited functionality of tools designed to interpret information, many aspects not preservable
- \- Cannot be applied to all types of objects

Strategies for Logical Preservation

Data Recovery, Data Archeology

- Analysis of bit-stream to interpret data, digital forensics
- \+ Probably only approach to recover "lost" information
- \- No guarantee that it works
- \- Without sufficient documentation close to "guessing"
- \- Extremely high costs per object
- \- Hard to estimate on whether it may be successful for a given object

Logical Preservation

Summary

- Many preservation actions, subtle variations
- Changing object, environment
- Loss upon migration / emulation
- Decision of what to preserve → **Significant Properties**!
- How to detect/document what you lost?
- Range of strategies available, none is perfect
- Combination of strategies
- Which solution is best? → **Preservation Planning!**
- No solution forever -> DP is a process!

Why do we need Digital Preservation

3 levels of threat / preservation

1. Bit rot – physical preservation / bit preservation
   Physical Layer: how to keep the 0's and 1's
2. Object formats – logical preservation
   Logical Layer: how to remain able to open a file, run a program
3. Authenticity, interpretability – semantic preservation
   Semantic Layer: how to ensure we can understand/interpret data correctly

- How? What can we do?

Semantic preservation

- Threats at semantic level
  - meaning of terms change: city names, …
  - measurement scales, sensor sensitivity, …change
  - interpretation of facts change: alcohol levels, …
  - data context-dependent: re-calibration/sensor drift
- Rather long-term, but **subtle to notice!**
- Consider context of objects
  - purpose, setting, limitations, cultural context, related objects, …

Semantic preservation

- Approaches / solutions:
  - Semantic enrichment
  - Metadata
  - Migration at semantic level
  - Documentation of context
  - Tracing of metadata
  - Document intended meaning / interpretation

From Data to Processes

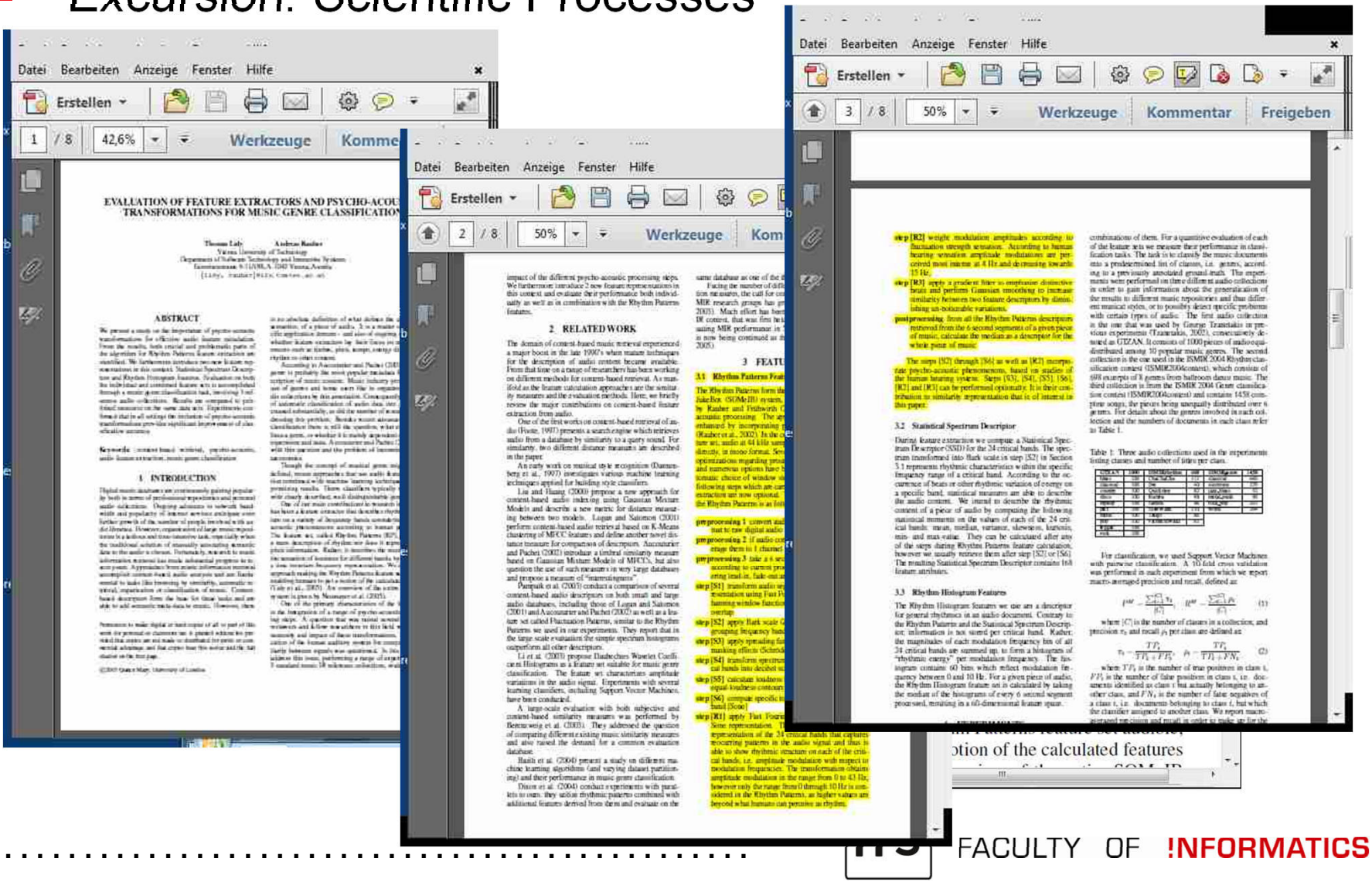
- Assume we know how to preserve data - **Is this sufficient?**
- Preserving data: Data Management Plans
  - describing data and context: provenance, authenticity, representation information,...
  - range of (ambiguous) definitions of context
  - But: mostly not actionable, not enforcable,...
  - BUT: data are (just) results of processes!
- Processes may be needed to
  - verify data
  - understand provenance
  - re-use process on new data
  - integrate data over time
- **Process curation instead of data curation!**

Outline

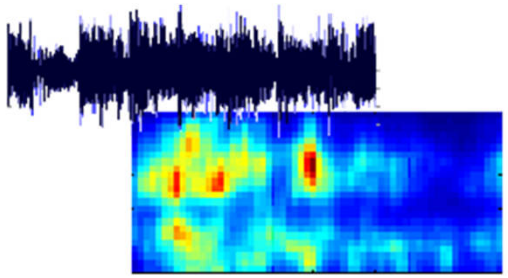
- Data-driven Science
- Digital Preservation
- From Data to Processes
- Technical solution: Process Capture
- Contractual solution: Holistic ESCROW
- How to evaluate?
- Data Curation and the Role of Digital Libraries

From Data to Processes

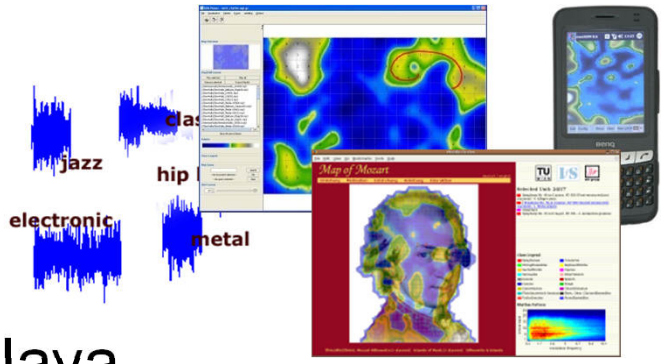
- Excursion: Scientific Processes



From Data to Processes

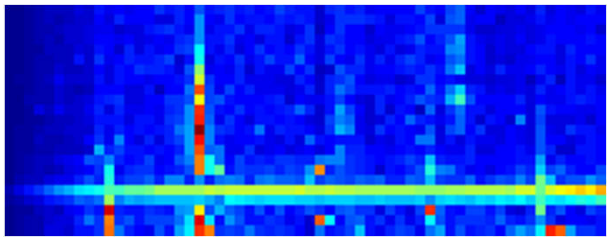


- Rhythm Pattern Feature Set
  - extracts numeric descriptors from audio
  - basically 2 Fourier Transforms
  - some psycho-acoustic modelling
  - some filters (gaussian, gradient) to make features more robust
- Used for
  - music genre classification
  - clustering of music by similarity
  - retrieval
- Implemented first in Matlab, then in Java
  - both publicly available on website
  - same same but different...

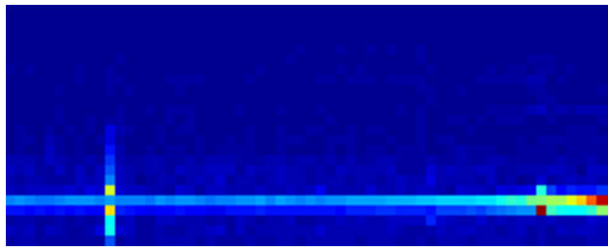


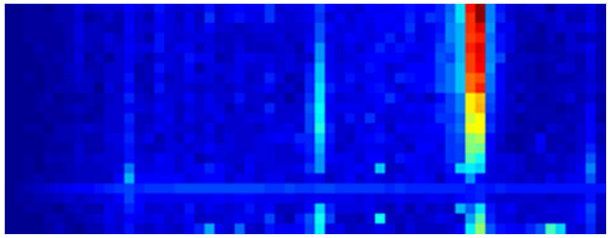
From Data to Processes

- Excursion: scientific processes

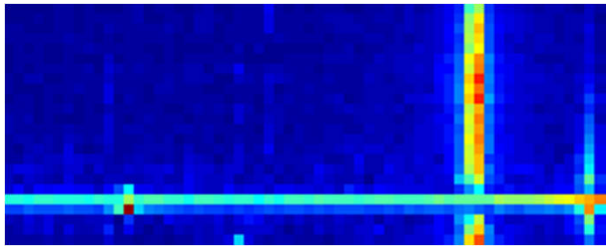


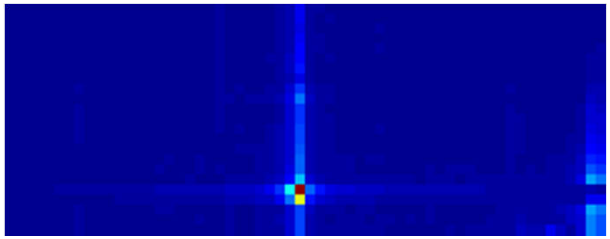
set1_freq440Hz_Am11.0Hz



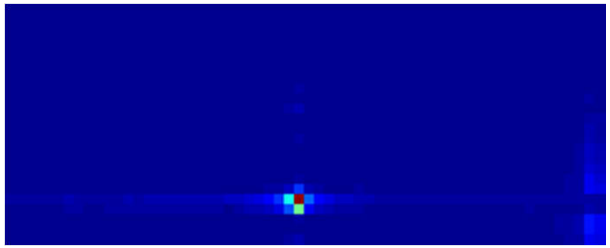


set1_freq440Hz_Am12.0Hz





set1_freq440Hz_Am05.5Hz

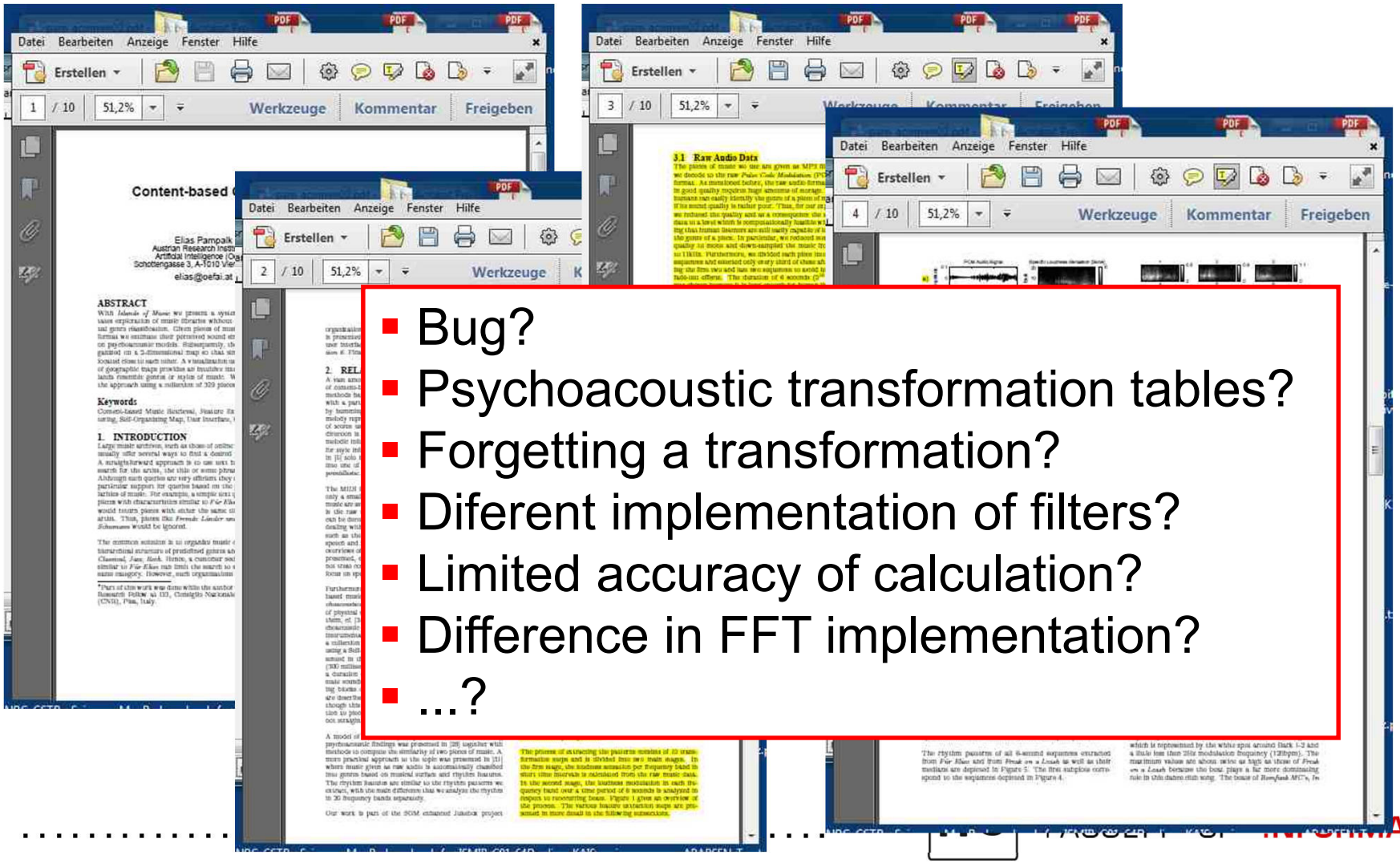


Java

Matlab

From Data to Processes

- Excursion: Scientific Processes



From Data to Processes

- Processes are important to understand data!
- Processes include
  - sensor capture (type, A/D conversion, calibration, operating conditions)
  - data (pre)processing: filtering, transformation
  - data integration: sources, transformations, treatment of missing values, outlier detection, ...
  - data analysis: tools, parameters, determinism
  - human operator activities
  - external services, web services
- End-to-end chain of activities underlying scientific experimentation
- Data as (interim) results

From Data to Processes

- Different disciplines of science, different means of validation
  - formal / proof
  - discourse
  - experimental evidence
- Many ICT-driven research areas experiment-driven
- How good are we in terms of repeatability/verifyability?
- Can we re-use earlier studies? verify code? share data?
- Need to ensure better procedures .....
  .....to support better science!

From Data to Processes

- How to curate processes?
  - how to capture and describe them?
  - what about proprietary elements?
  - how to evaluate if curation/re-activation is successful?

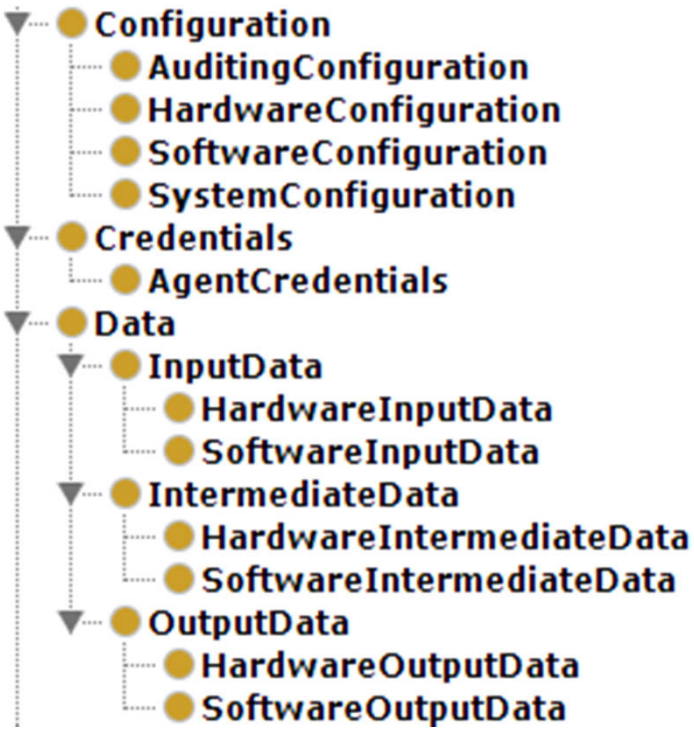
    (sig-props for processes and how to measure)

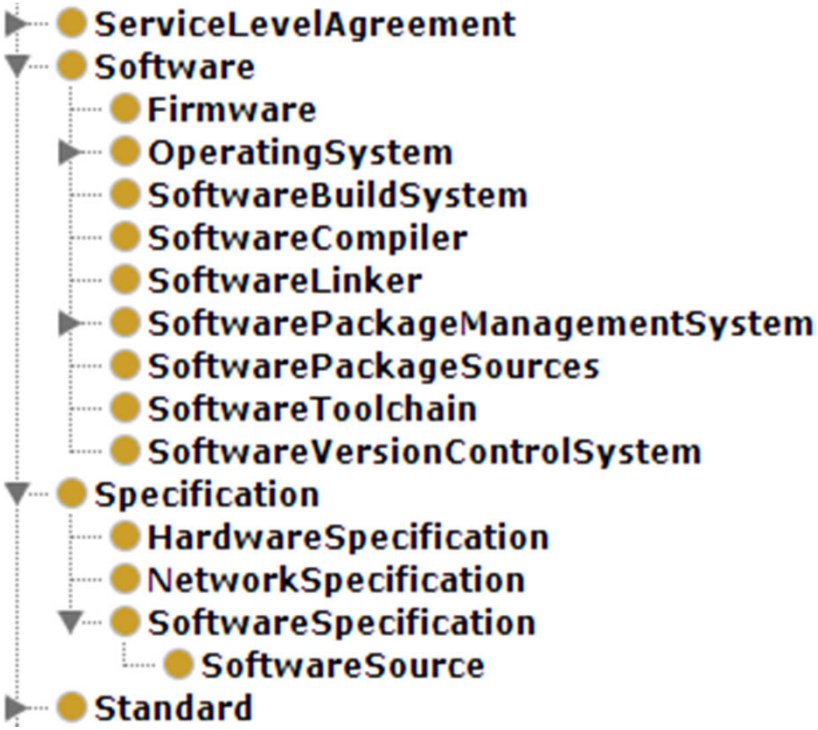
Outline

- Data-driven Science
- Digital Preservation
- From Data to Processes
- Technical solution: Process Capture
- Contractual solution: Holistic ESCROW
- How to evaluate?
- Data Curation and the Role of Digital Libraries

Process Capture

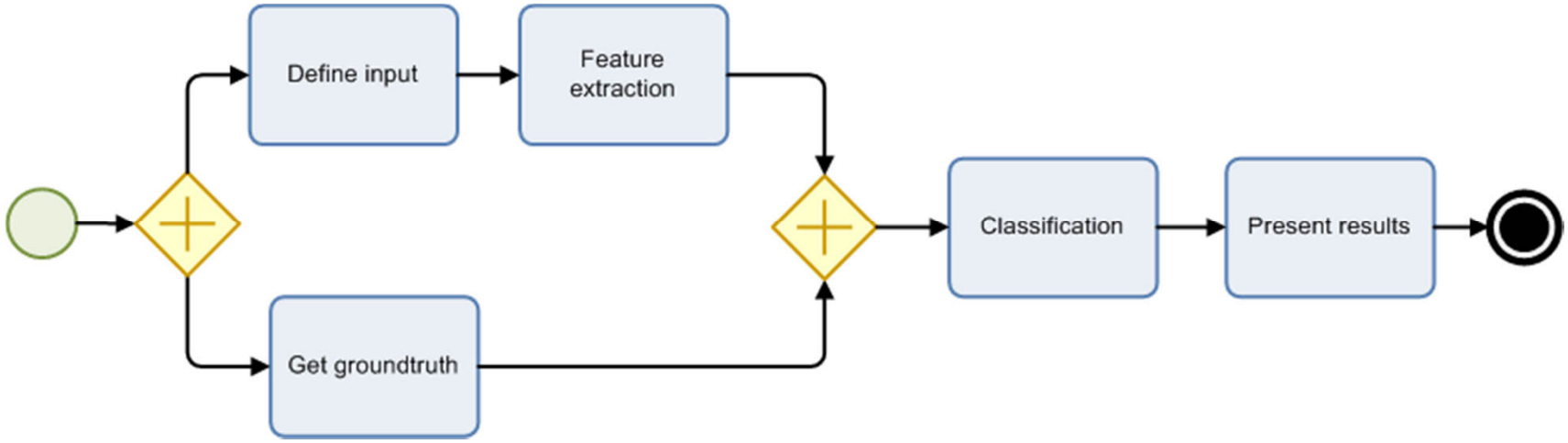
- Need to establish what forms part of a process:
  - analyzing process documentation
  - establishing context of process, relationships between elements
  - monitoring of process activities
- Capture and describe this in a context model





Process Capture

Example: Music Classification Process



- Input: Music (e.g. MP3 format)
- Input: Training data, i.e. music with genre labels
- Output: Classification of Music, e.g. into genres
- Intermediate steps
  - extract numeric description (features) from music
  - combine features with ground truth into specific file format

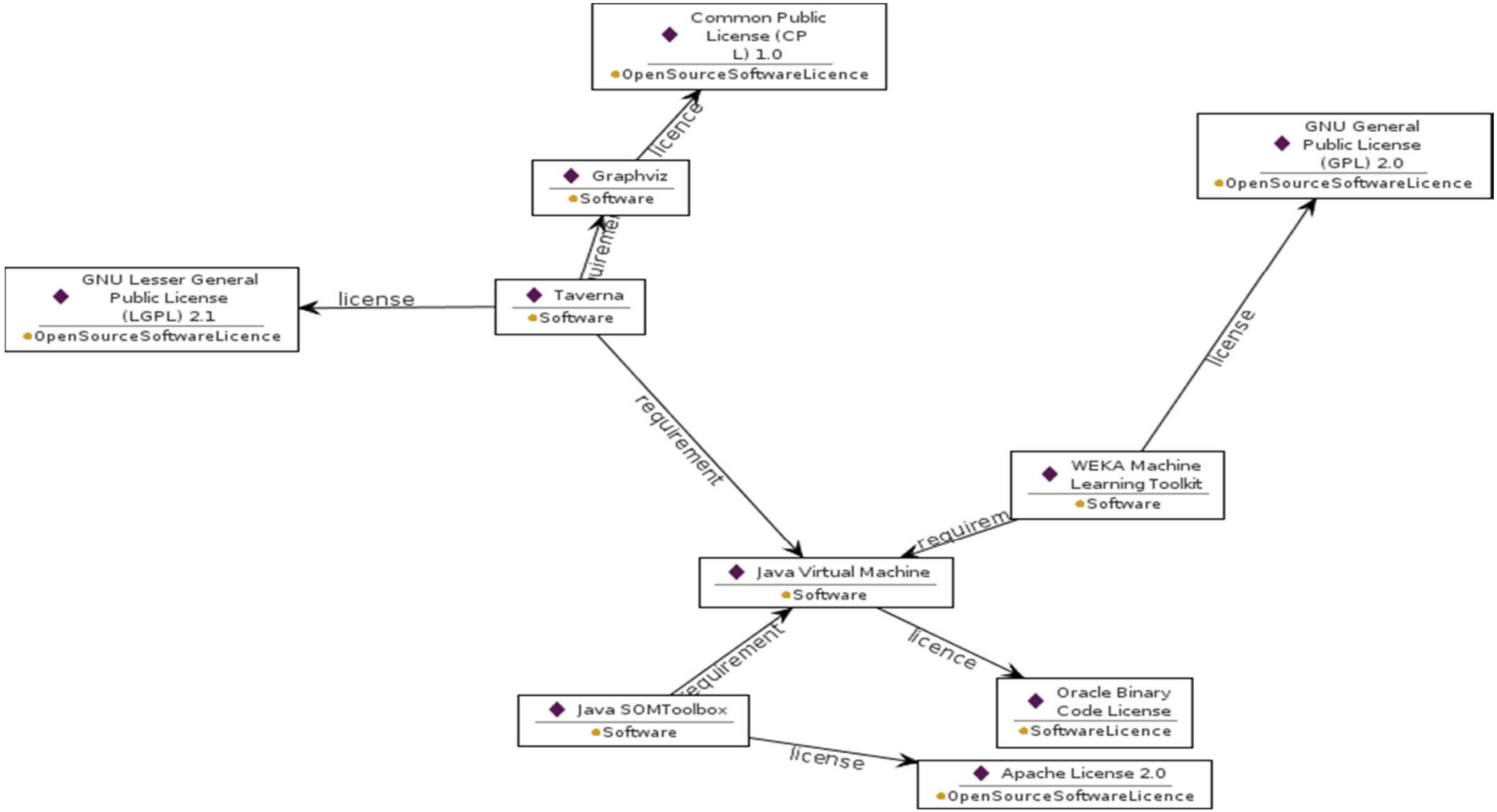
Process Capture

- Similar to Representation Information Networks, but extended to capture broader process context

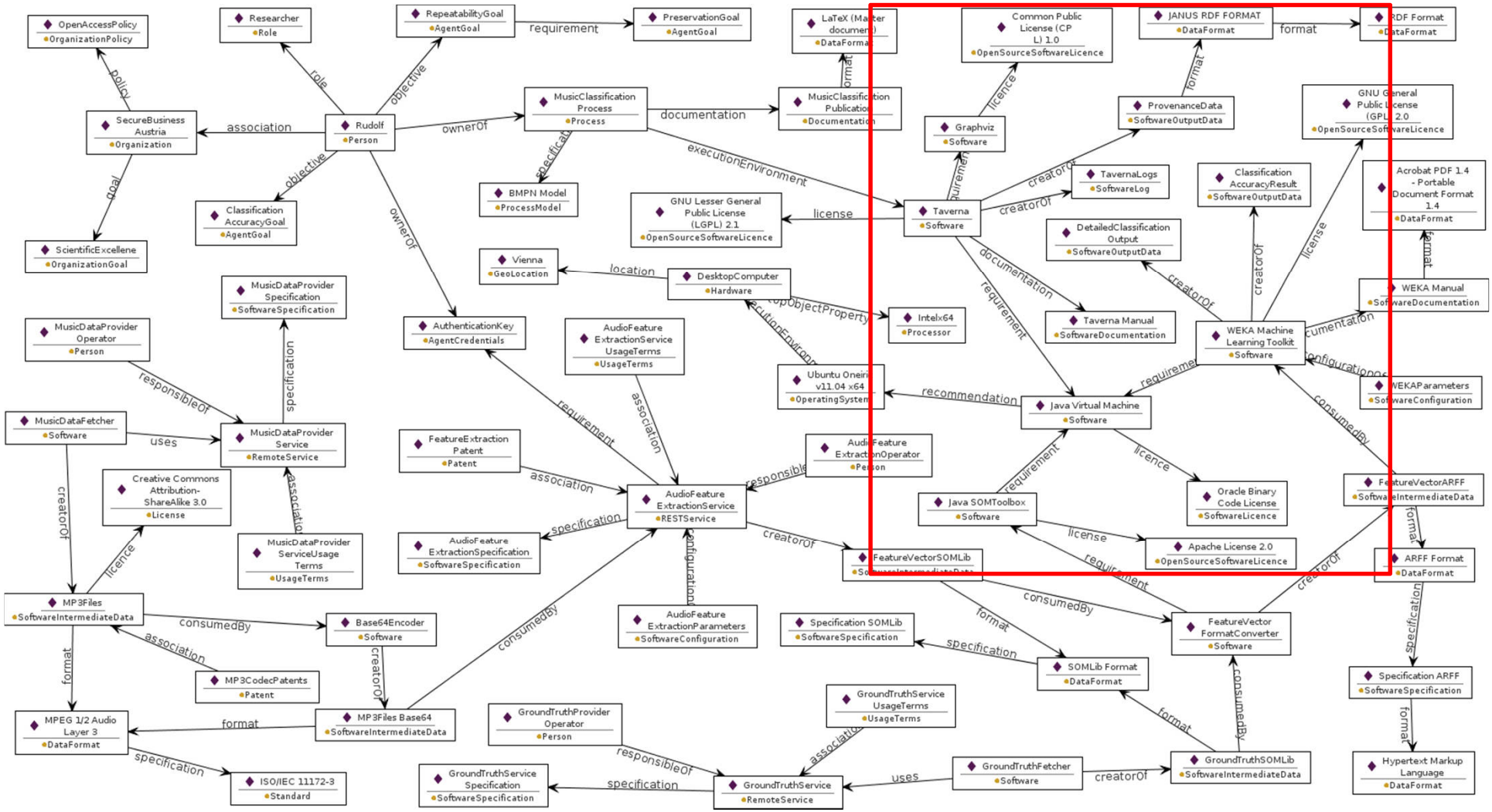
- Derived via top-down and bottom-up approach
  - enterprise architecture frameworks such as ZACHMAN
  - existing taxonomies, such as PREMIS
  - derived from scenarios developed by project partners
  - intellectual property rights, data analysis, software escrow, multimedia services, …

Process Capture

- Software setup can be automatically detected in OS with software packages (e.g. Linux); allows detection of licenses



Process Capture



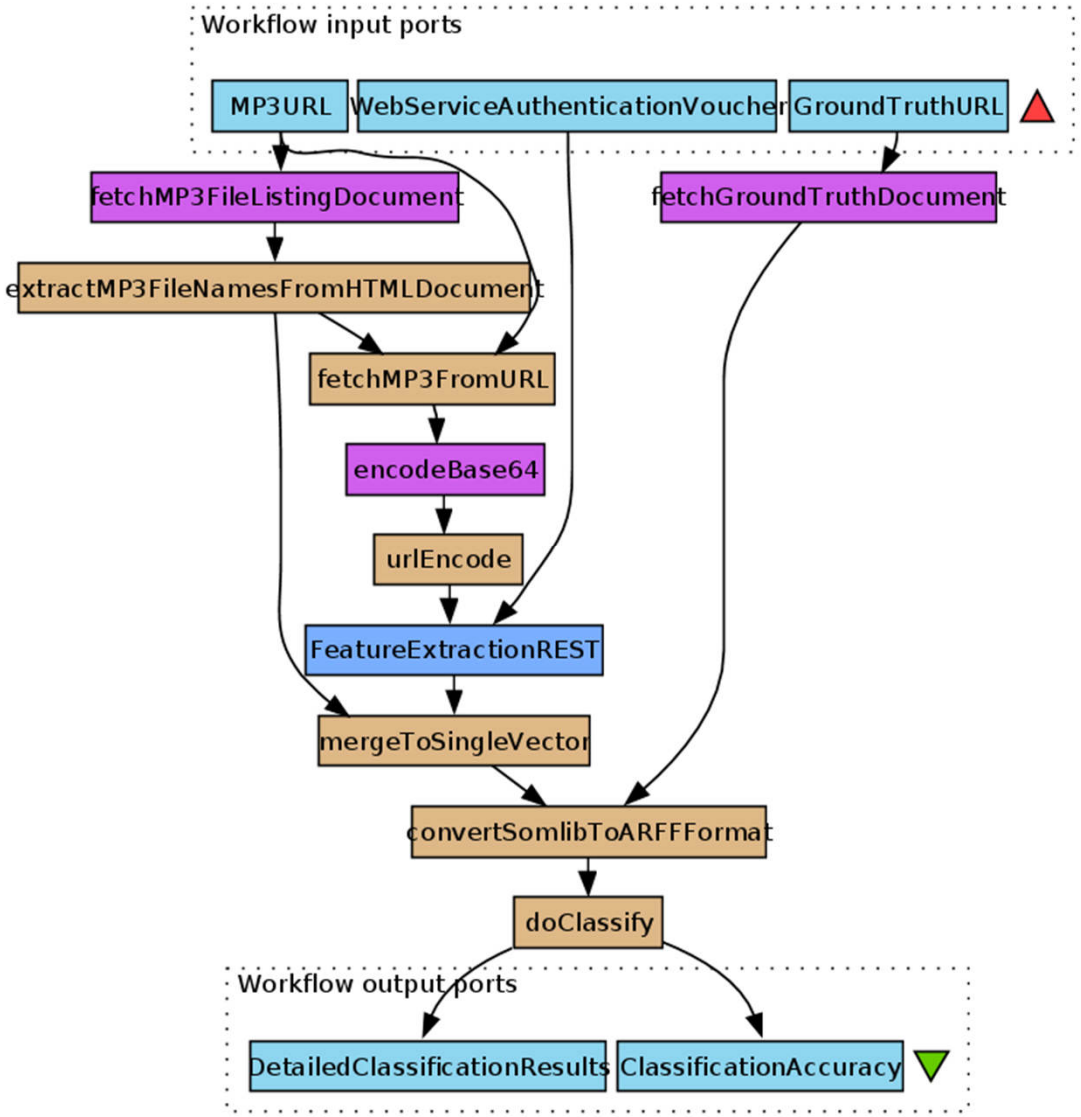
Process Capture

Recommendation

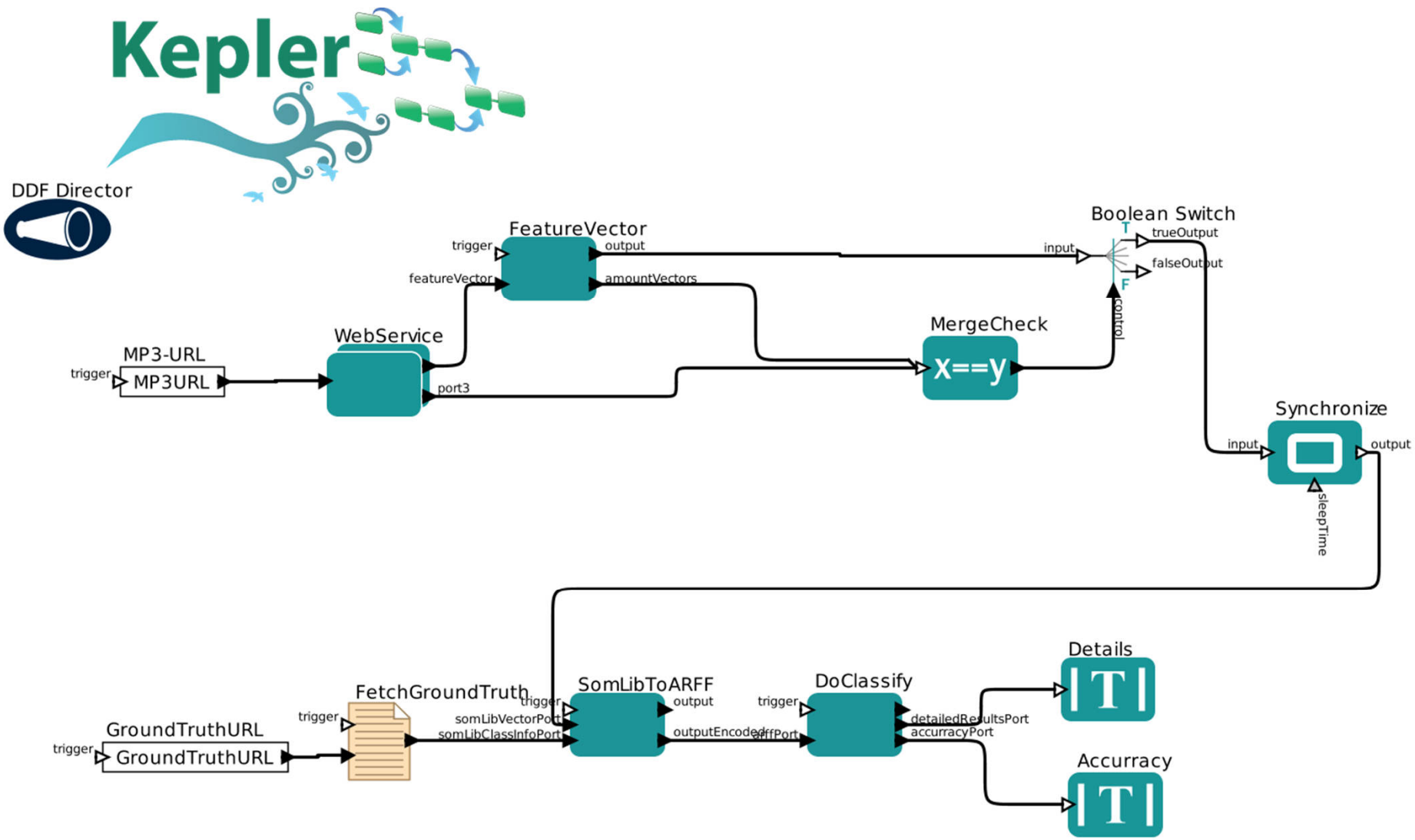
- Establish solidly documented research processes ("lab books")
- Rely on "preservable components" only
- Stability
- Example: modelling in Taverna, Kepler, Activity,...

Process Capture

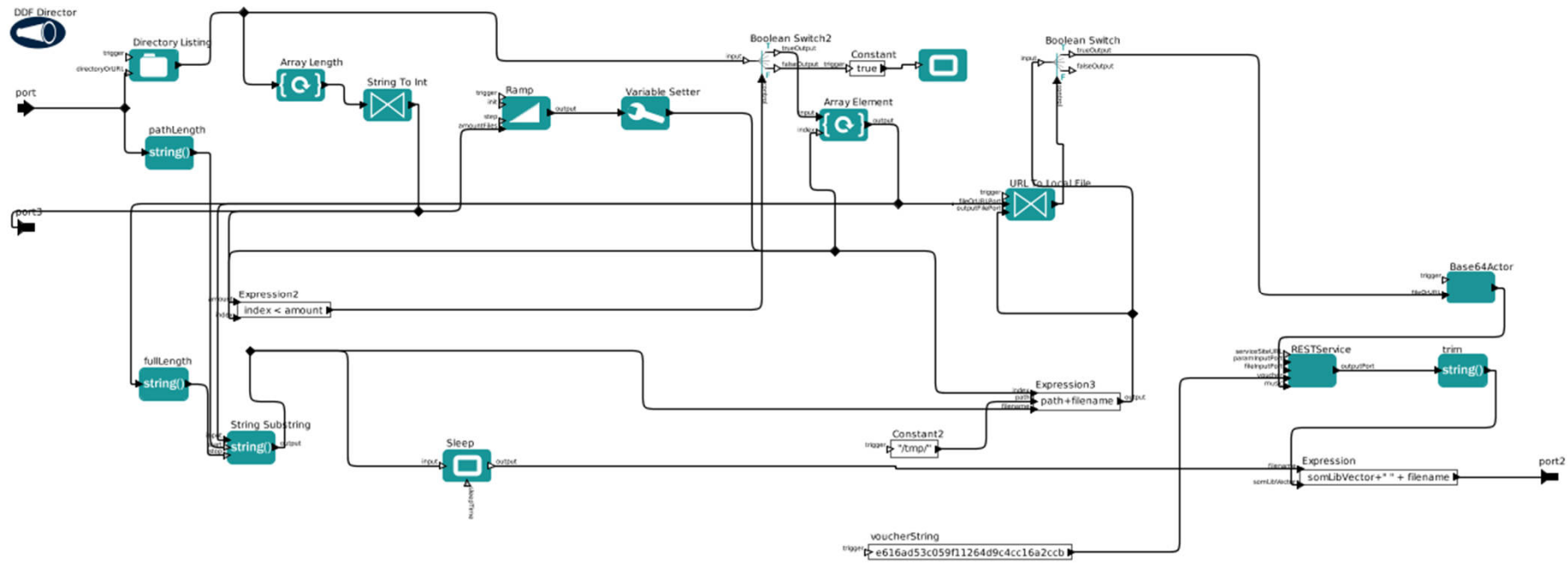
Taverna



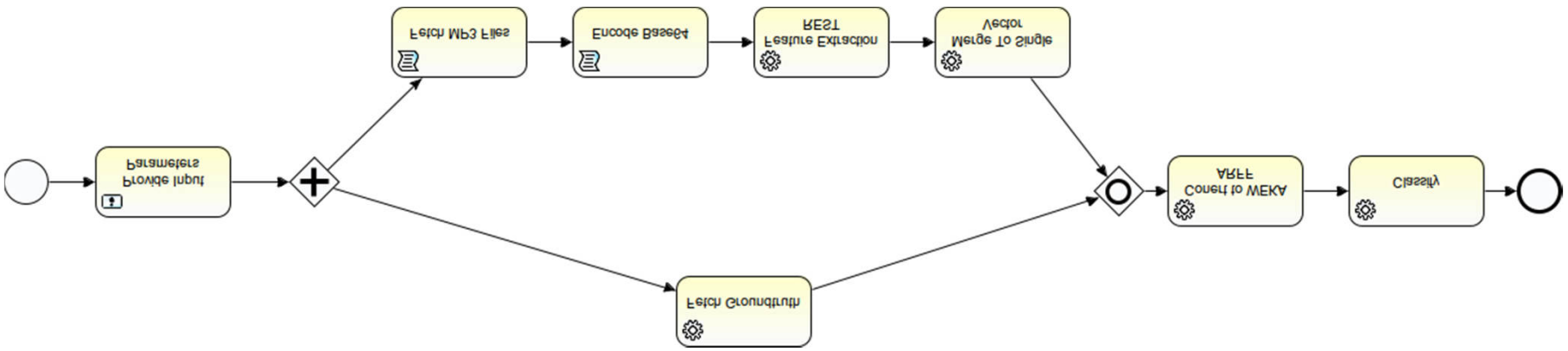
Process Capture



Process Capture



Activiti

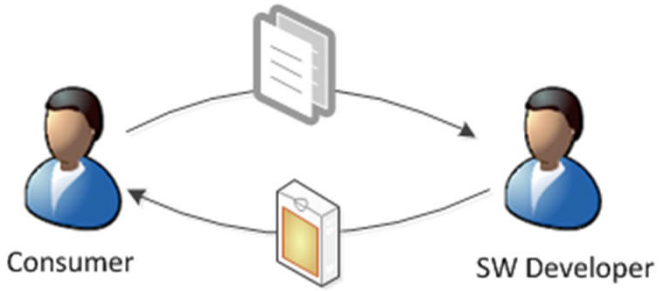

Parameters
Provide Input
Fetch MP3 Files
Encode Base64
REST
Feature Extraction
Merge To Single
Vector
Fetch Groundtruth
ARFF
Convert to WEKA
Classify

Outline

- Data-driven Science
- Digital Preservation
- From Data to Processes
- Technical solution: Process Capture
- Contractual solution: Holistic ESCROW
- How to evaluate?
- Data Curation and the Role of Digital Libraries

ESCROW

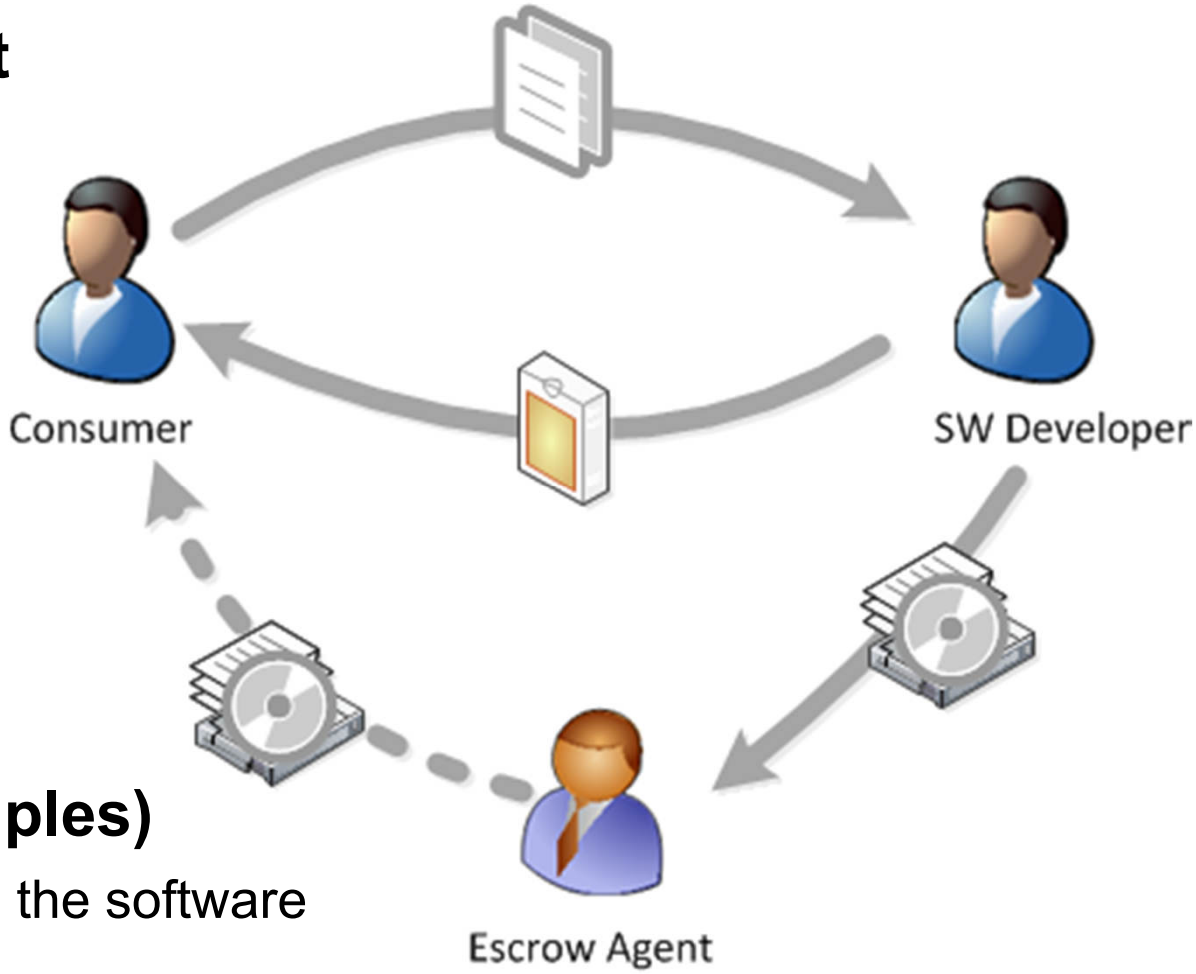
- Assume we have the process described
- Assume we have all modules captured
- What to do about
  - 3rd-party tools?
  - external services?
  - HW modules with proprietary configurations?
- ESCROW agreements for trusted third-party deposit
- Mitigating risks such as
  - Financial standing of the vendor
  - Sale of the vendor
  - Maintenance of the system
  - Loss of cooperation



ESCROW

ESCROW Agreement

- Consumer
- Developer
- ESCROW Agent



Release events (examples)

- Failure to support/maintain the software
- Insufficient maintenance
- Insolvency/ bankruptcy of SW-developer

ESCROW

Obligations and Motivations

- Consumer
  - Use of the software
  - Protection of investment
- SW Developer
  - Provides object code to consumer
  - Commits source code to Escrow Agent
  - Obligations to the consumer as well as the Software Escrow Agent
- Escrow Agent
  - Verification of deposit material
  - Deposit of source code
  - In case of agreed circumstances release of the source code to the consumer

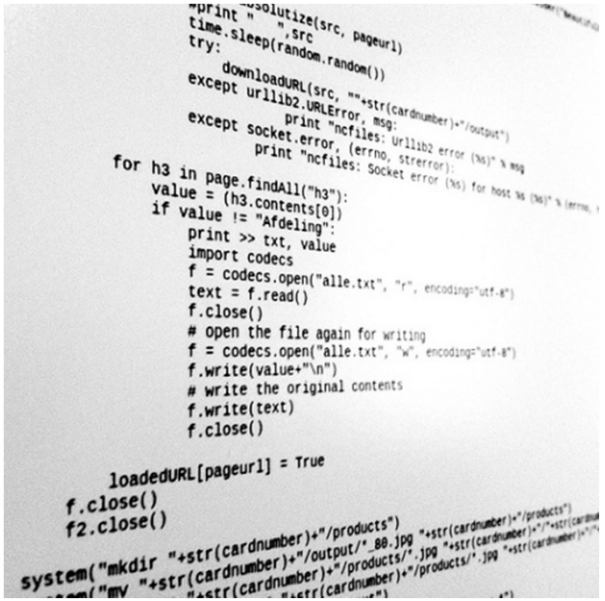
ESCROW

Motivation

- Consumer
  - risk mitigation strategy
    - maintenance and support
    - protection of business
  - hedging of investments
  - verification of source code
- SW-Developer
  - evidence of copyright
  - warranty claim
  - confidence-building measure
  - payment from customer

Holistic ESCROW

Problems/Challenges



- Read error storage media
- Source code incomplete
- Build environment not deposited
- Configuration not available
- Instructions missing
- Test data missing
- Documentation insufficient
- Licences not included (Development- & Build-Environment, Libraries)
- Deposited material is not up to date
- ...

Holistic ESCROW

Requirements

- Completeness
  - source code only part of a software
  - without additional information almost impossible to understand, analyse, use and change the source code
- Quality
  - up to date
  - maintainability
- Need to verify as far as possible automatically

Holistic ESCROW

Artifacts of a software product

Holistic ESCROW

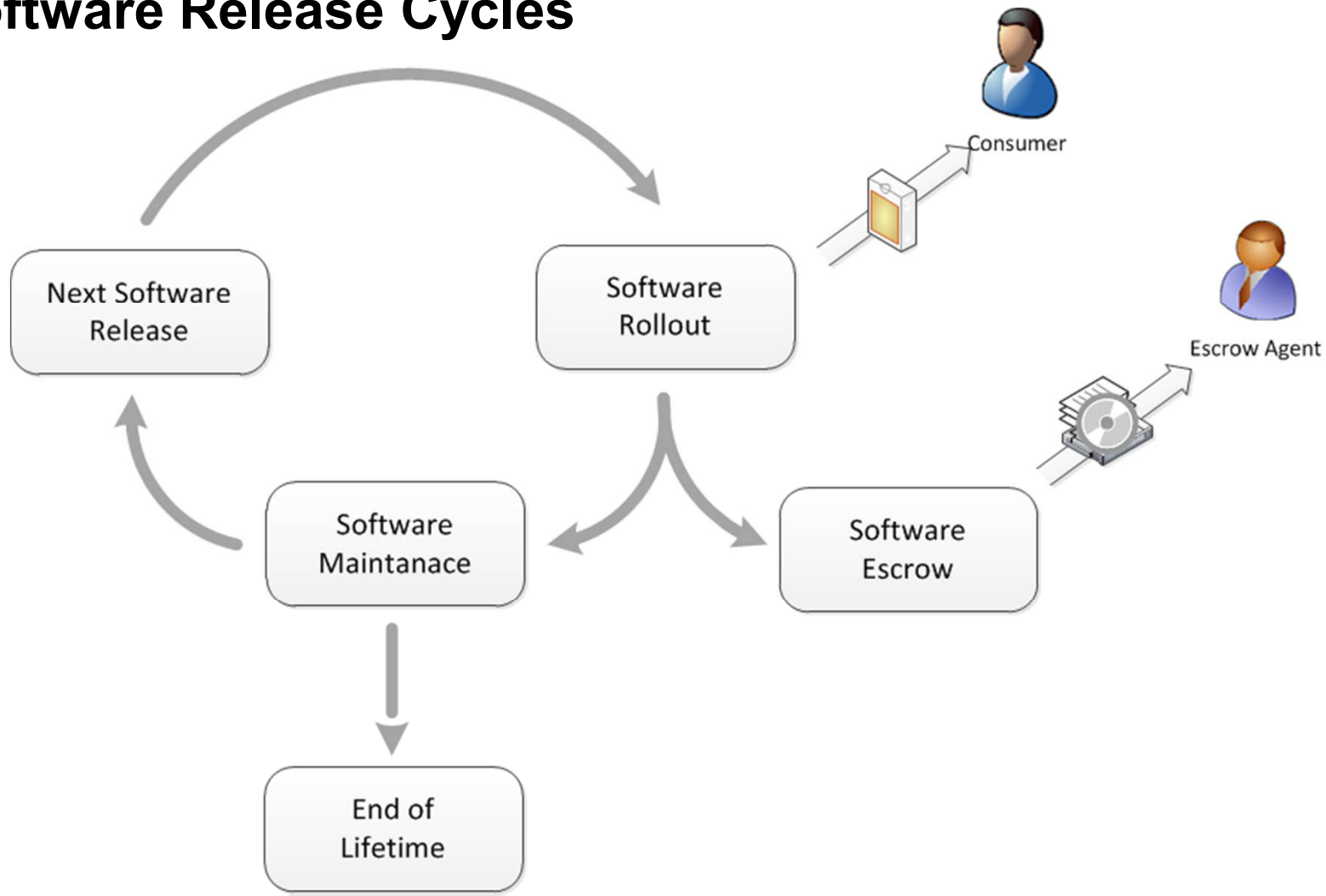
ESCROW Audits

- Verification in literature
  - Standard or basic
    Readable and virus free material, check list
  - Full verification
    compile of source code
  - Bespoke verification
    negotiable (ensuring to meet specification, manual review)

- Costs vs benefit

Holistic ESCROW

Software Release Cycles



Escrow Audit

ESCROW Audit

- Completeness of deposit material
  - Source code alone is not sufficient

- Quality measurements
  - Documentation
  - Understandability
  - Testability
  - Modifiability

- Automated analysis of deposit material, manual cross-checking

Outline

- Data-driven Science
- Digital Preservation
- From Data to Processes
- Technical solution: Process Capture
- Contractual solution: Holistic ESCROW
- How to evaluate?
- Data Curation and the Role of Digital Libraries

How to evaluate?

- Properties of a digital object that are considered significant and as such have to be preserved
- Examples
  - image width, colour depth
  - page breaks, font, character encoding
  - relative speed
  - ...
- Preservation action should preserve the important significant properties
- How to apply this to processes? easy?

How to evaluate?

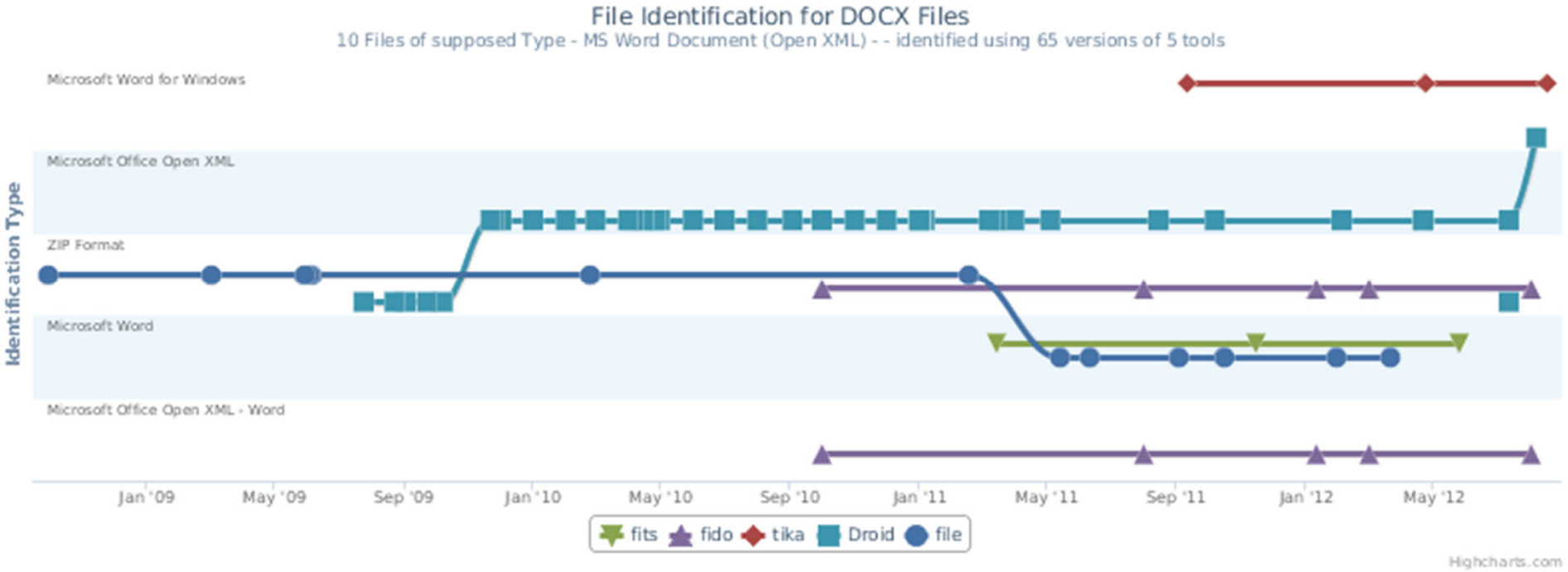
Challenges in Property Verification

- Identifying properties
- Identifying metrics
- Identifying where to measure/extract properties
  - properties of file format vs. rendered properties
  - characteristics of view path!
- Scaling: millions of objects
- Reliability of property extractors

How to evaluate?

Challenges in Property Verification (2)

- Reliability of property extractors: example: file format



Tarrant, David and Carr, Leslie (2012)
LDS3: Applying Digital Preservation Principals to Linked Data Systems.
9th Intl. Conf. on Digital Preservation (iPres2012), Toronto, Canada

How to evaluate?

Further Problems with dynamic and interactive content:

- To get reproducible results the digital object has to follow a deterministic behaviour:
  - what are the factors that influence the objects behaviour?
- Continuous rendering of objects:
  - when should object properties be extracted?
  - where can properties be extracted from the running system?

How to evaluate?

Deterministic behaviour:

- View path has to be constant to compare behaviour
- Input has to be constant
  - macros
  - remote access
  - “hardware” (read input on hardware level on original system, apply on hardware-layer of emulator)
- External factors that influence deterministic behaviour have to be constant (e.g. date/time, network activity, random number seed)
- Not every object‘s behaviour can be made deterministic! (or not with justifiable effort)

How to evaluate?

How to extract significant properties:

- Not from the object...
- ... but from the environment (object is rendered by/in the environment)
- Environment has to support extraction
- Properties have the dimension time (e.g. frames/second, cycles per second, number of file access operations per minute)
- Properties change over time (e.g. frames/second min, average, max)

How to evaluate?

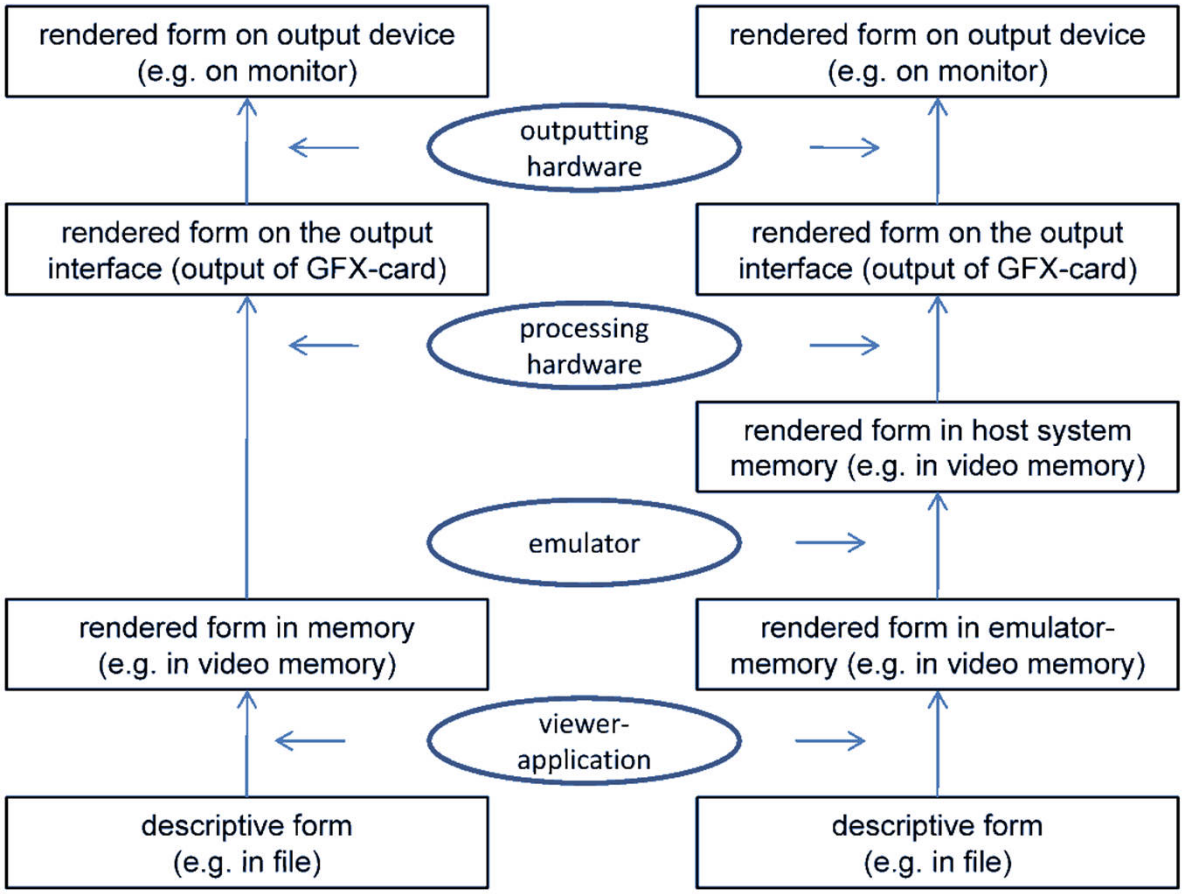
When to extract properties:

- Not every state in an object's rendering process is significant
- Depending on the object
  - **target state:** only one state after initially rendering the object or after applying a certain series of input events (e.g. rendering a static object)
  - **sequence of states:** only certain states after certain events (e.g. web site after each click on a link)
  - **continuous stream:** every rendered state of the object is important (e.g. video game, sound stream)

How to evaluate?

Where to extract properties:

- Rendered form of a digital object exists on various levels in a system:



How to evaluate?

Where to extract properties:

- Descriptive form
  - before rendering, useful for migration, no change in emulation
- Rendered form in memory (original system)
  - both original system and emulator memory of host system
- Rendered form in memory (host system)
  - exists only on host system, no comparison possible
- Rendered form on output interface
  - signal analyzed by comparator (e.g. digital video signal, analog audio signal)
- Rendered form on display device
  - influence of display device options (e.g. brightness settings)
- **Note**: identical results on one level may require different results on levels below!

How to evaluate?

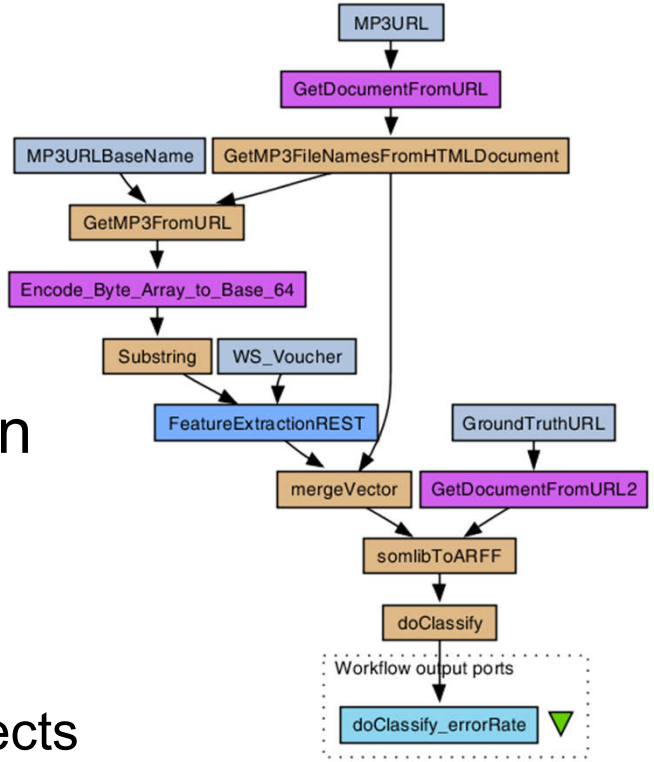
Where to extract properties:

- Depending on the tests where to extract:
  - screenshot level after applying input: is the emulation working correctly ?
  - comparison of output: is the transformation from emulated system to host system working correctly ?
- Depending on the original system:
  - screenshot only possible if not a single process system like video game device
  - output signal only possible if output device is not part of system (e.g. built-in speakers)
  - after output device always possible, additional factors influencing the signal

How to evaluate?

Example:

- Music Workflow is running in an emulated environment (virtual machine, actual emulator)



- Points of interest in Workflow from an external view would be
  - GetDocumentFromURL: interface with expects input
  - GetDocumentFromURL2: interface expects input
  - GetMP3FromURL
  - Workflow output ports: interface provides output

How to evaluate?

Validation Workflow:

1. Describe the original environment
2. Determine external events that influence the object's behavior -> make it deterministic
3. Decide on what level to compare the digital object
4. Recreate the environment (emulation, viewer, ...)
5. Apply standardized input to both environments
6. Extract significant properties
7. Compare the significant properties

Outline

- Data-driven Science
- Digital Preservation
- From Data to Processes
- Technical solution: Process Capture
- Contractual solution: Holistic ESCROW
- How to evaluate?
- Data Curation and the Role of Digital Libraries

Role of DLs

- Digital Libraries / Repository Systems need to move beyond papers:
  - data as part of publications
  - processes as part of data
- DLs as eScience repository systems
- Advise on Data (Process) Management Plans
- Establishing (virtual) centers of expertise, services
- Some disciplines ahead of others
  - genome database, meteorology, crystallography, high energy physics, ...
  - emergent networks in the humanities (DARIAH)
  - less established in many core ICT domains
- (Are we a proper science?)

Role of DLs

- Complexity of challenges underlying research data and process preservation
- Does every researcher need to become a DP specialist?
  - support for researchers, embedding of information specialists
- Training and education
  - need for information specialists to assist researchers
  - understand DP, ICT and research domain
  - DL as interdisciplinary community to drive curricula development
- What do we need to teach?
  - metadata schemas, curation, concepts of information
  - database theory, programming, security, sensor and embedded systems

Role of DLs

- Global cooperation sharing R&D as well as operational costs
  - shared DP workflows (SCAPE)
  - approaches for process preservation (Timbus, Workflow4ever)
  - virtual center of excellence (APARSEN, DPE)
  - understanding costs (4C)
- Discipline-specific & national repository networks
- Joint training and educational programs
- Cooperation in research, but also...
- ... cooperation in operations

Conclusions

- We are entering a new era of science - need to adapt!
- Data turning into key enabler of scientific discovery
- Big data and long-tail
- Data only result of processes
- Process curation
  - capturing & documentation
  - establish well-documented workflows to ease curation
  - holistic ESCROW for third-party modules
  - solid procedures for evaluation
- Huge challenges for data-intensive / eScience / 4th paradigm / ... research
- Key challenge & opportunity for DL community

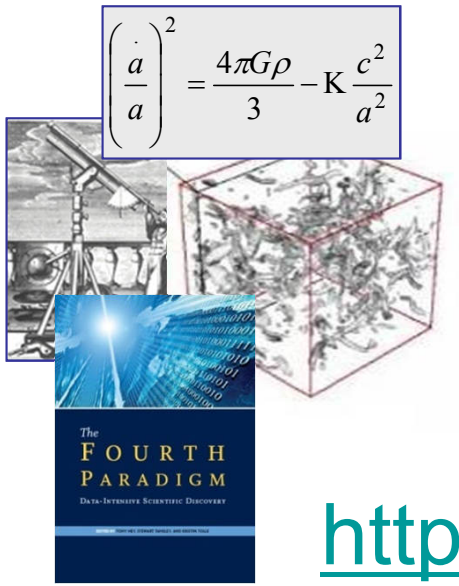
Thanks

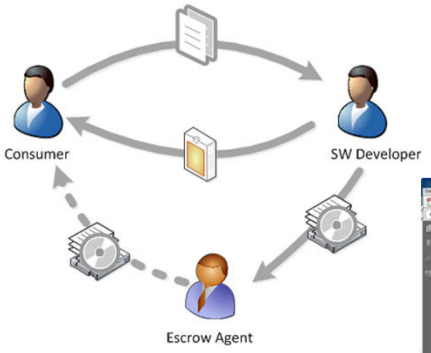
- Research projects
  - SCAPE
  - TIMBUS
  - APARSEN

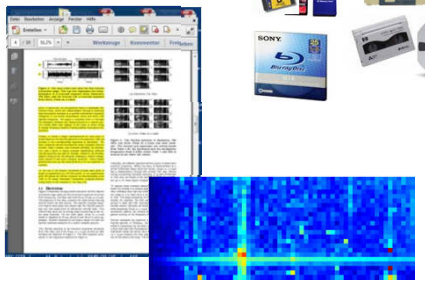
Thanks

- Meetings and discussion fora
  - NSF Workshop on Data Quality, Sep. 10/11 2012, Arlington, USA
    http://datacuration.web.unc.edu/
  - ANA-DP: Aligning National Approaches to Digital Preservation
    May 23-25 2011, Tallin, Estonia
    http://www.educopia.org/events/ANADP
    + follow-up events at iPRES 2012; IDCC1013; ANADP2, Fall 2013
  - Research Challenges in DP:
    - Dagstuhl seminar, July 19-23 2010
      http://drops.dagstuhl.de/opus/volltexte/2010/2899/pdf/10291.Report.2899.pdf
    - iPRES ORC Workshop, Oct. 2 2012
      http://digitalpreservationchallenges.wordpress.com/
    - DP? wiki
      http://sokrates.ifs.tuwien.ac.at/wiki/index.php/Main_Page

Thank you!







http://www.ifs.tuwien.ac.at/imp