Preservation planning 3

What to measure and how
Monitoring and Scalability
Multi-criteria decision making

April 22, 2013

Christoph Becker

Vienna University of Technology
www.ifs.tuwien.ac.at/~becker

Preservation Planning: Question 1

- **How can we select the optimal preservation action for a given setting?**
  - What are the drivers and constraints on the decision space?
  - What are the goals and objectives?
  - What are the factors influencing the decision makers' preferences?
  - How can we model multiple competing objectives and requirements?
  - How should we evaluate software components?

Preservation Planning: Question 2

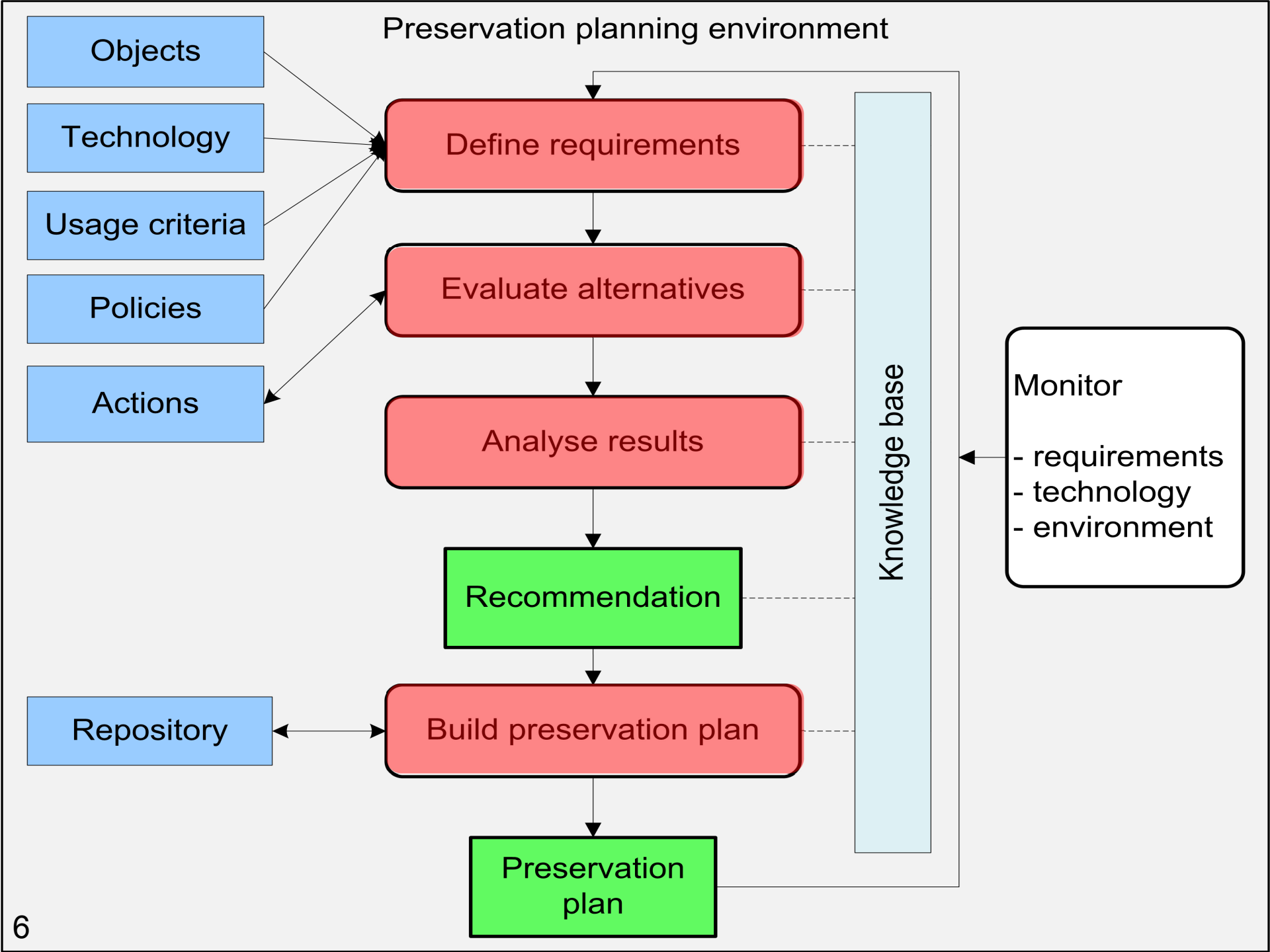
- **How can we ensure trustworthy preservation planning?**
  - What are the requirements on trust that need to be addressed?
  - What decision steps and evidence need to be documented?
  - What are the aspects that a plan needs to address, and what are the elements needed to cover them?
  - How can we ensure reliable evaluation procedures and repeatable evidence?

Preservation Planning: Question 3

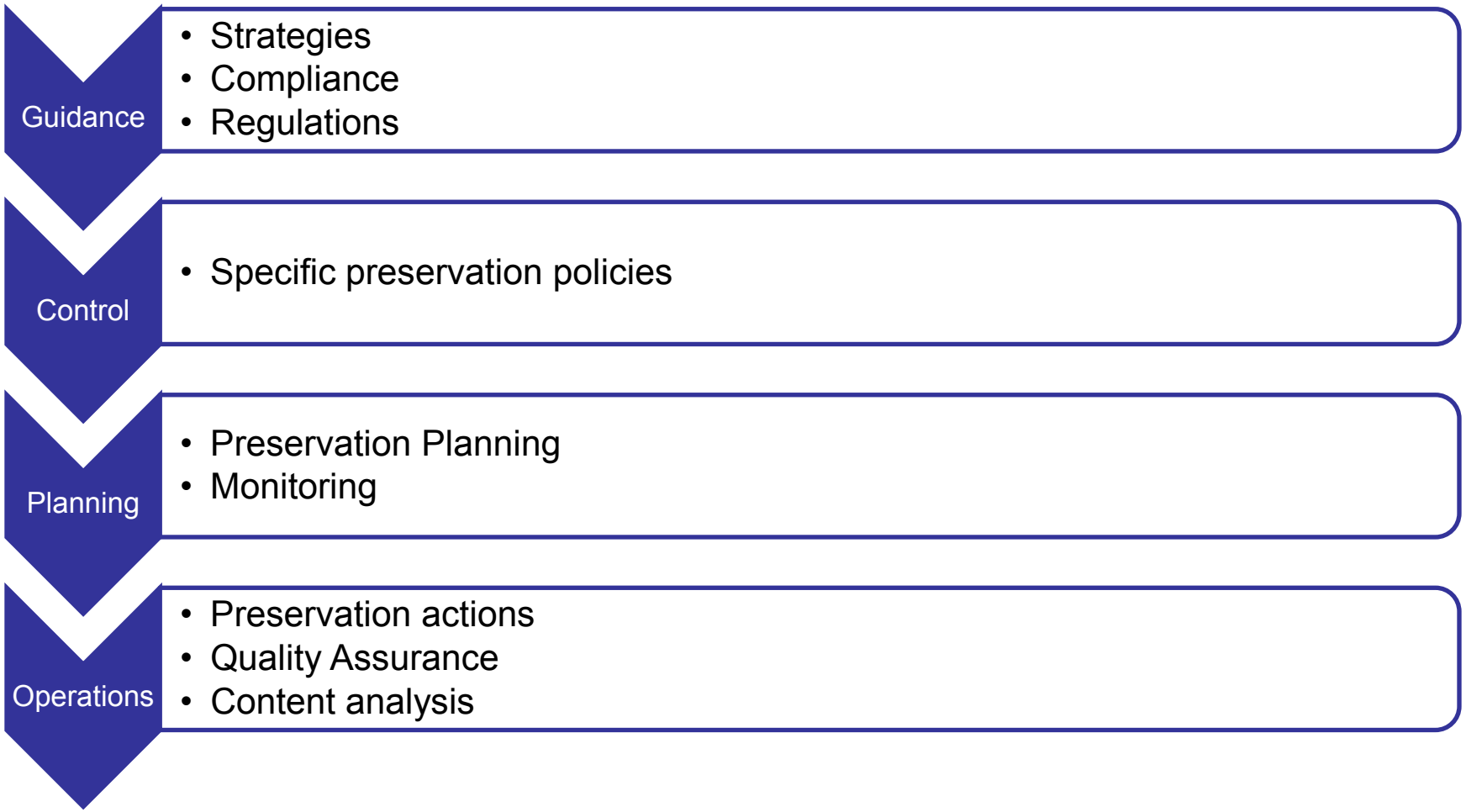
- **How can we ensure that decision processes scale up?**
  - How can we automate decision making?
  - How can we integrate continuous monitoring?
  - Which properties can be measured automatically, and how?
  - How can we create a controlled environment for observing the behaviour of components in a reproducible way?

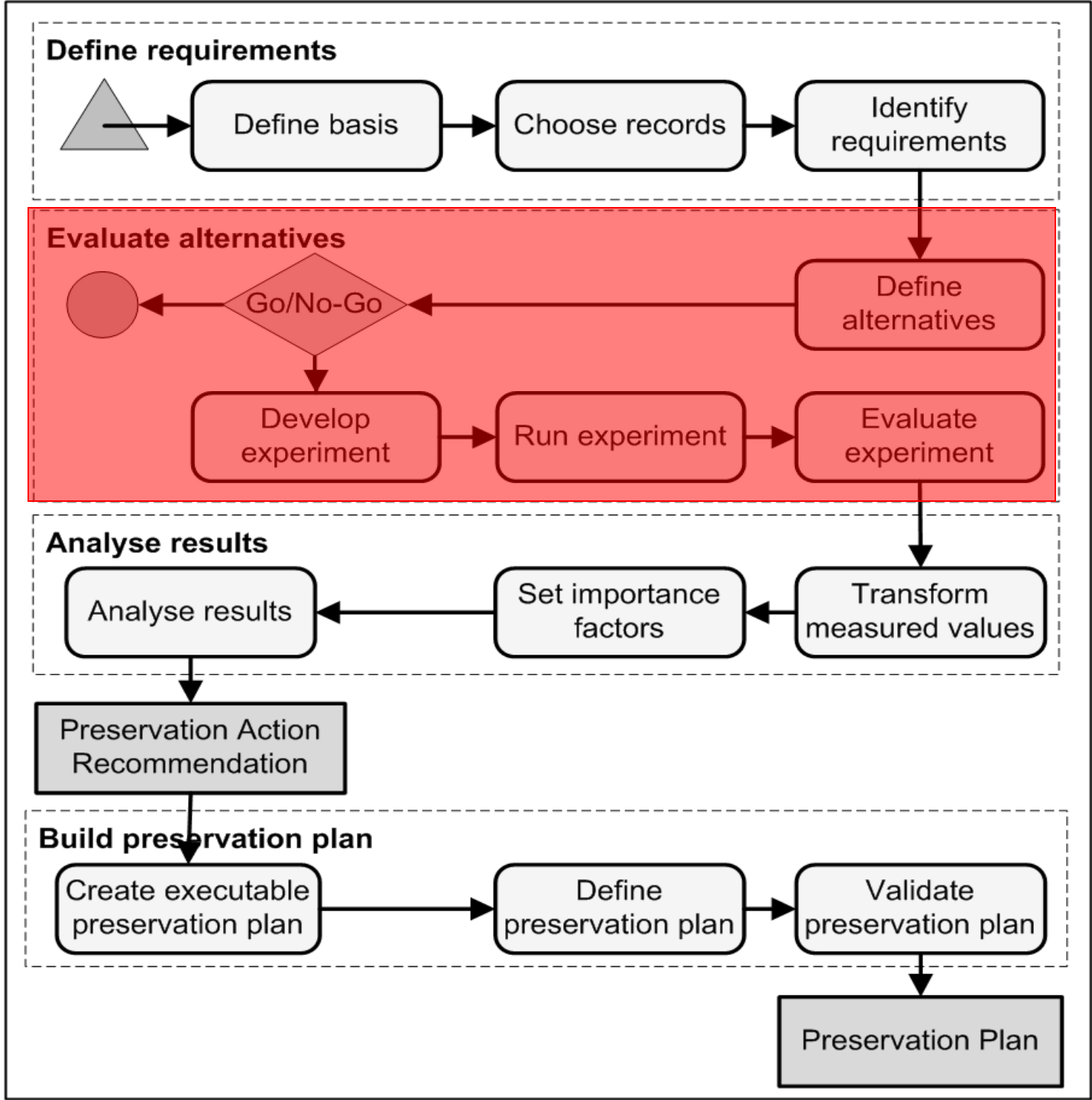
Preservation Planning: Key concepts

- Repeatable, standardized planning workflow
- A weighted hierarchy of objectives
  - Measurable criteria on the leaf level of the tree
  - Utility functions make criteria comparable
- Controlled experimentation on sample content
  - Evidence-based decision making
- Standardized structure for plan specification
  - Transparency and documentation
  - Comparability across scenarios
  - Integration with repository systems
- Planning tool Plato guides, validates, documents
  - www.ifs.tuwien.ac.at/dp/plato
- Automation: Reduce manual effort

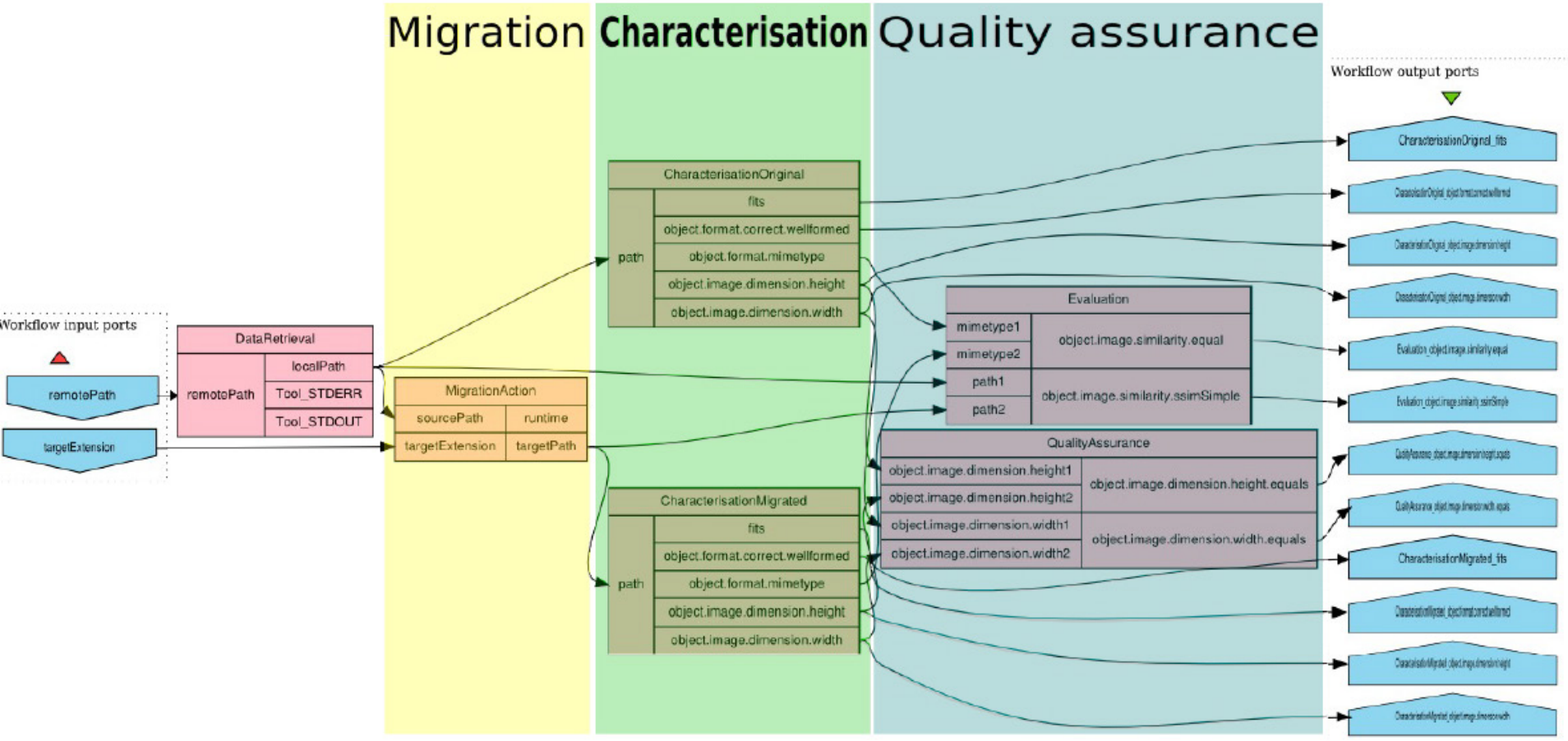

Preservation planning environment
Objects
Technology
Usage criteria
Policies
Actions
Define requirements
Evaluate alternatives
Analyse results
Recommendation
Build preservation plan
Preservation plan
Repository
Knowledge base
Monitor
- requirements
- technology
- environment

Levels of control



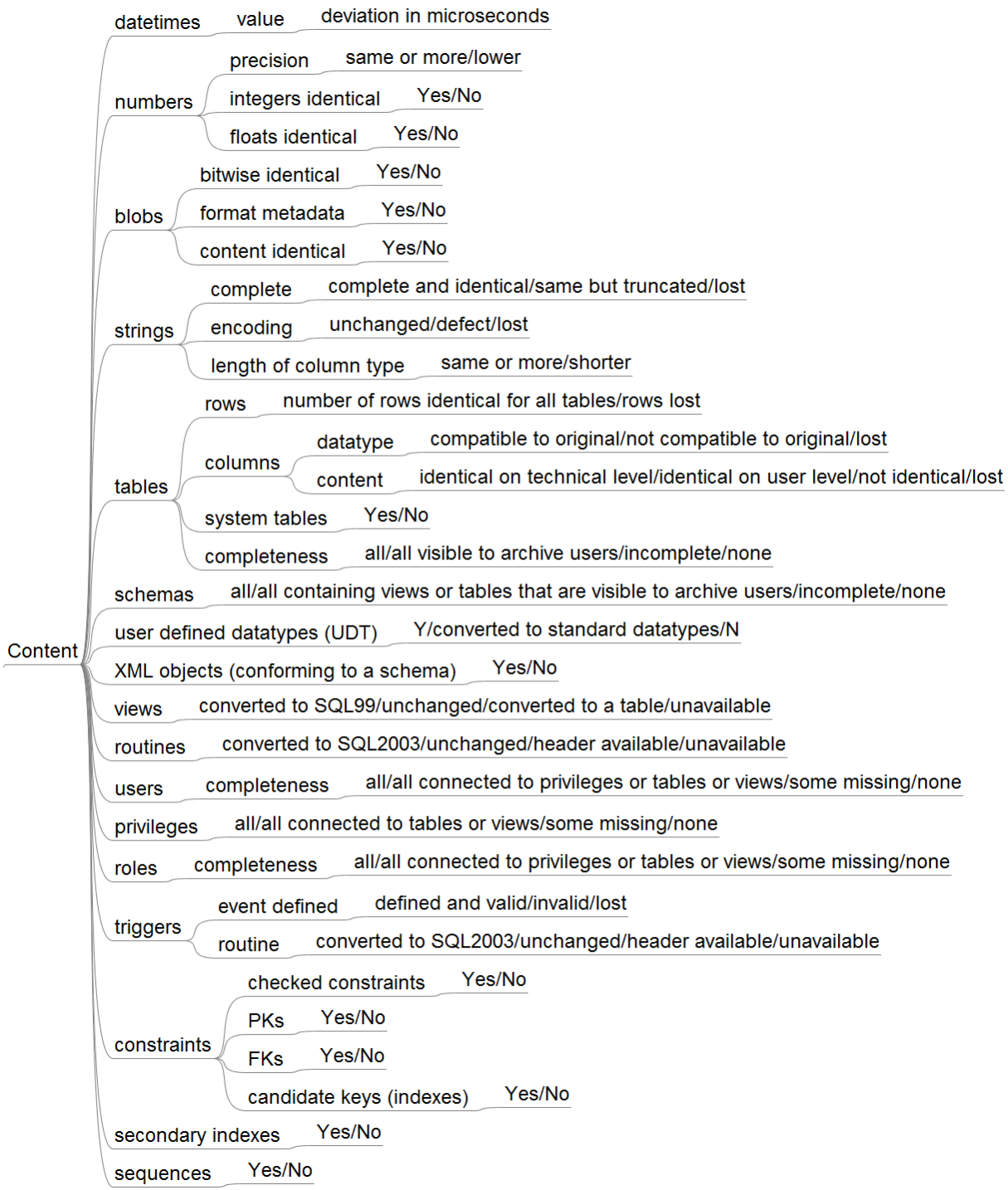

Define requirements
Define basis
Choose records
Identify requirements
Evaluate alternatives
Go/No-Go
Define alternatives
Develop experiment
Run experiment
Evaluate experiment
Analyse results
Analyse results
Set importance factors
Transform measured values
Preservation Action Recommendation
Build preservation plan
Create executable preservation plan
Define preservation plan
Validate preservation plan
Preservation Plan

A planning experiment



Database study

Content branch



Results: Weighted multiplication

Result-Tree with all Alternatives, Aggregation method: Weighted multiplication

Node	Results	
Requirements	Archive to XML:	3.88
	Keep original DB:	3.72
	CSV export:	0.00
Object characteristics	Archive to XML:	2.14
	Keep original DB:	2.32
	CSV export:	0.00
Content	Archive to XML:	2.38
	Keep original DB:	2.51
	CSV export:	0.00
appearance	Archive to XML:	1.14
	Keep original DB:	1.16
	CSV export:	1.06
context	Archive to XML:	1.23
	Keep original DB:	1.17
	CSV export:	0.88
behaviour	Archive to XML:	1.00
	Keep original DB:	1.12
	CSV export:	1.00
Format characteristics	Archive to XML:	1.36
	Keep original DB:	1.20
	CSV export:	1.13
Tool characteristics	Archive to XML:	1.34

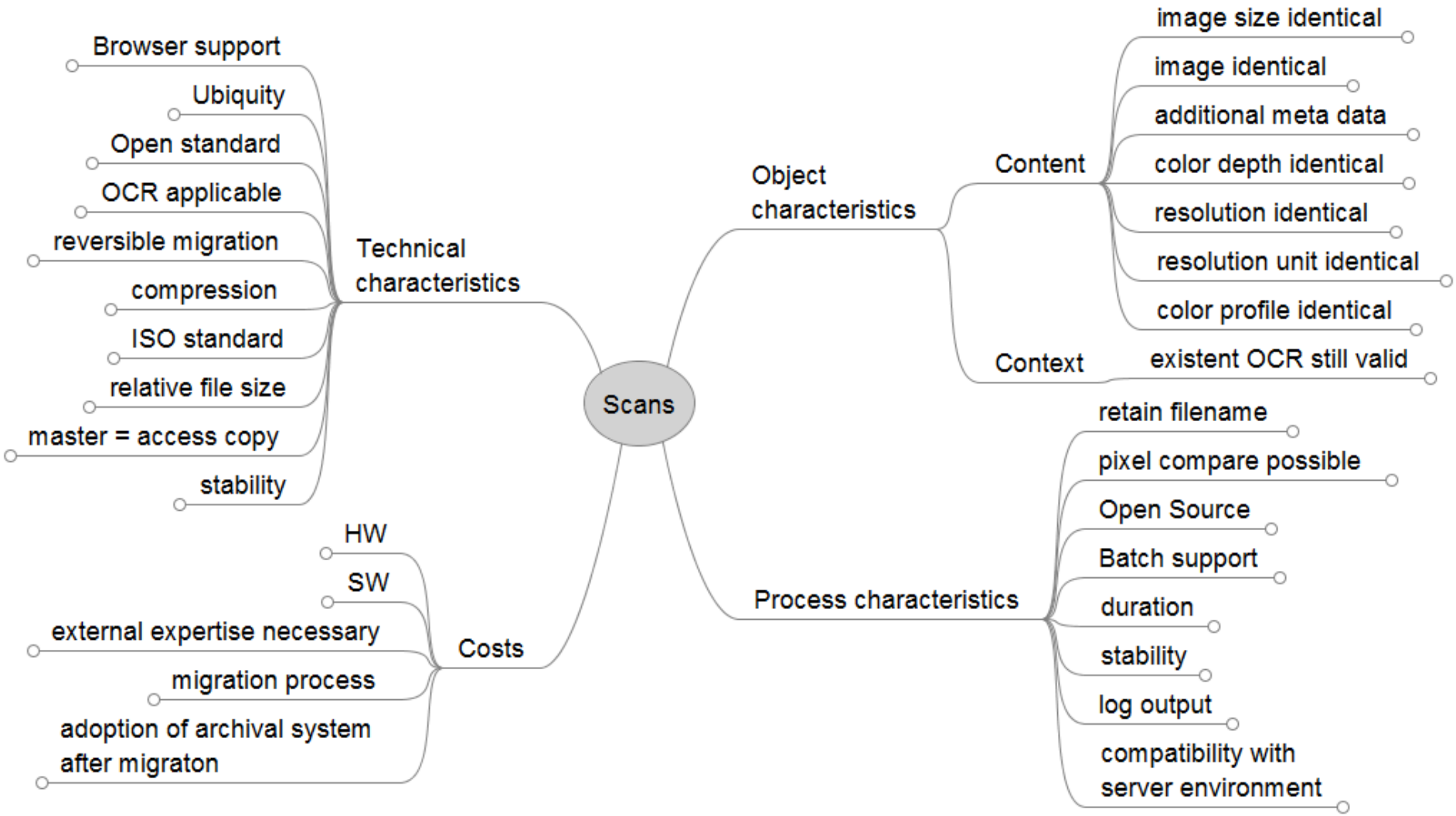
Results: Weighted sum

Result-Tree with all Alternatives, Aggregation method: Weighted sum.

This tree contains only strategies that do not have knock-out evaluation criteria; see above

Node	Results	
Requirements	Archive to XML:	4.25
	Keep original DB:	4.07
Object characteristics	Archive to XML:	2.40
	Keep original DB:	2.60
Content	Archive to XML:	2.63
	Keep original DB:	2.76
appearance	Archive to XML:	0.41
	Keep original DB:	0.45
context	Archive to XML:	0.69
	Keep original DB:	0.57
behaviour	Archive to XML:	0.07
	Keep original DB:	0.35
Format characteristics	Archive to XML:	0.95
	Keep original DB:	0.57

Scanned books requirements



Scanned books results

Results: Weighted multiplication

Result-Tree with all Alternatives, Aggregation method: Weighted multiplication

Node	Results	
Scans	Keep status quo:	4.50
	ImageMagick - TIFF to JP2:	3.71
	GraphicsMagick - TIFF to JP2:	0.00
	Kakadu - TIFF to JP2:	3.68
	GeoJasper - TIFF to JP2:	3.65
Object characteristics	Keep status quo:	1.50
	ImageMagick - TIFF to JP2:	1.38
	GraphicsMagick - TIFF to JP2:	0.00
	Kakadu - TIFF to JP2:	1.38
	GeoJasper - TIFF to JP2:	1.38
Content	Keep status quo:	2.24
	ImageMagick - TIFF to JP2:	1.63
	GraphicsMagick - TIFF to JP2:	0.00
	Kakadu - TIFF to JP2:	1.63
	GeoJasper - TIFF to JP2:	1.63
image size identical	Keep status quo:	1.19
	ImageMagick - TIFF to JP2:	1.19
	GraphicsMagick - TIFF to JP2:	1.19
	Kakadu - TIFF to JP2:	1.19
	GeoJasper - TIFF to JP2:	1.19
image identical	Keep status quo:	1.19
	ImageMagick - TIFF to JP2:	1.19
	GraphicsMagick - TIFF to JP2:	0.00
	Kakadu - TIFF to JP2:	1.19
	GeoJasper - TIFF to JP2:	1.19
additional meta data	Keep status quo:	1.19
	ImageMagick - TIFF to JP2:	1.08
	GraphicsMagick - TIFF to JP2:	1.08
	Kakadu - TIFF to JP2:	1.08
	GeoJasper - TIFF to JP2:	1.08
color depth identical	Keep status quo:	1.19
	ImageMagick - TIFF to JP2:	1.19
	GraphicsMagick - TIFF to JP2:	1.19
	Kakadu - TIFF to JP2:	1.19
	GeoJasper - TIFF to JP2:	1.19

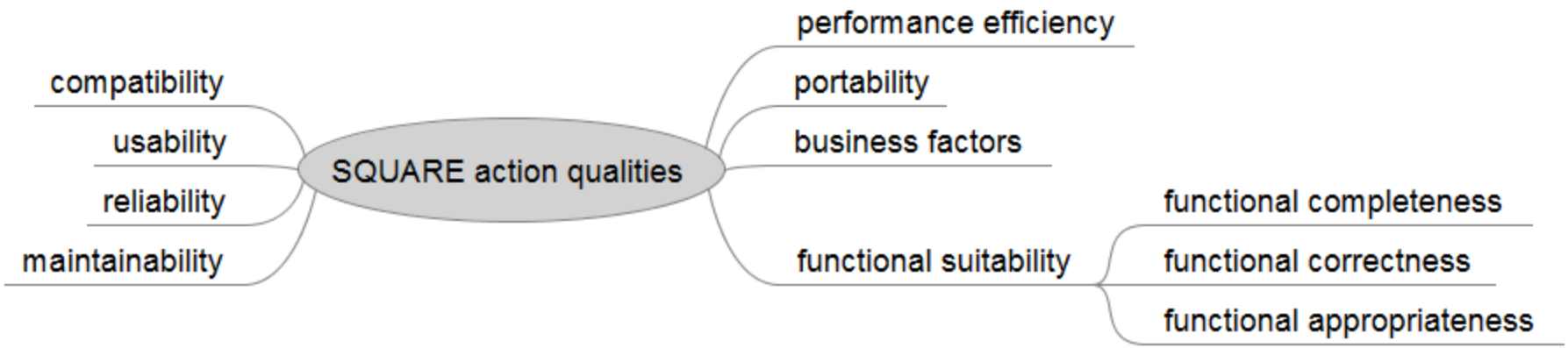
Four cases, three solutions: Scanned images

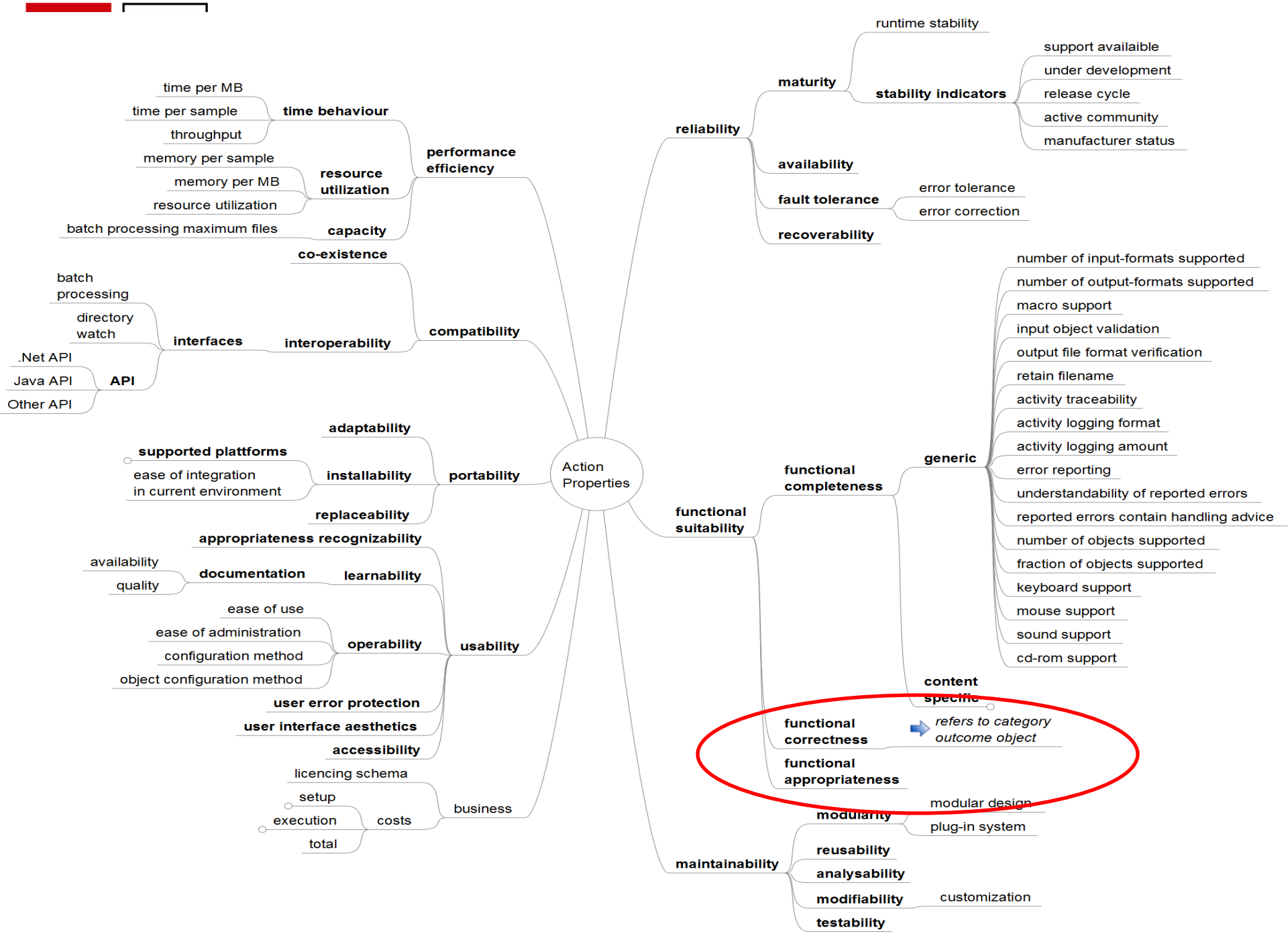
- Bavarian State Library, 72TB TIFF6: *Leave and monitor*
- British Library, 80TB TIFF5: *Migrate to JP2 (ImageMagick)*
- Royal Library of Denmark, ~10.000 aerial photographs in TIFF6: *Leave and monitor*
- State and University Library Denmark, scanned yearbooks in GIF: *Migrate to TIFF 6*

Scenario	Chosen action	Main reasons
72 TB scanned book pages in TIFF6	Leave unchanged and monitor	Color profile complications, lack of JP2 browser support, Process costs
80 TB scanned newspapers in TIFF5	Migrate to JP2	Storage costs, Standardization
Aerial photographs in TIFF6	Leave unchanged and monitor	Lack of JP2 browser support, Process costs

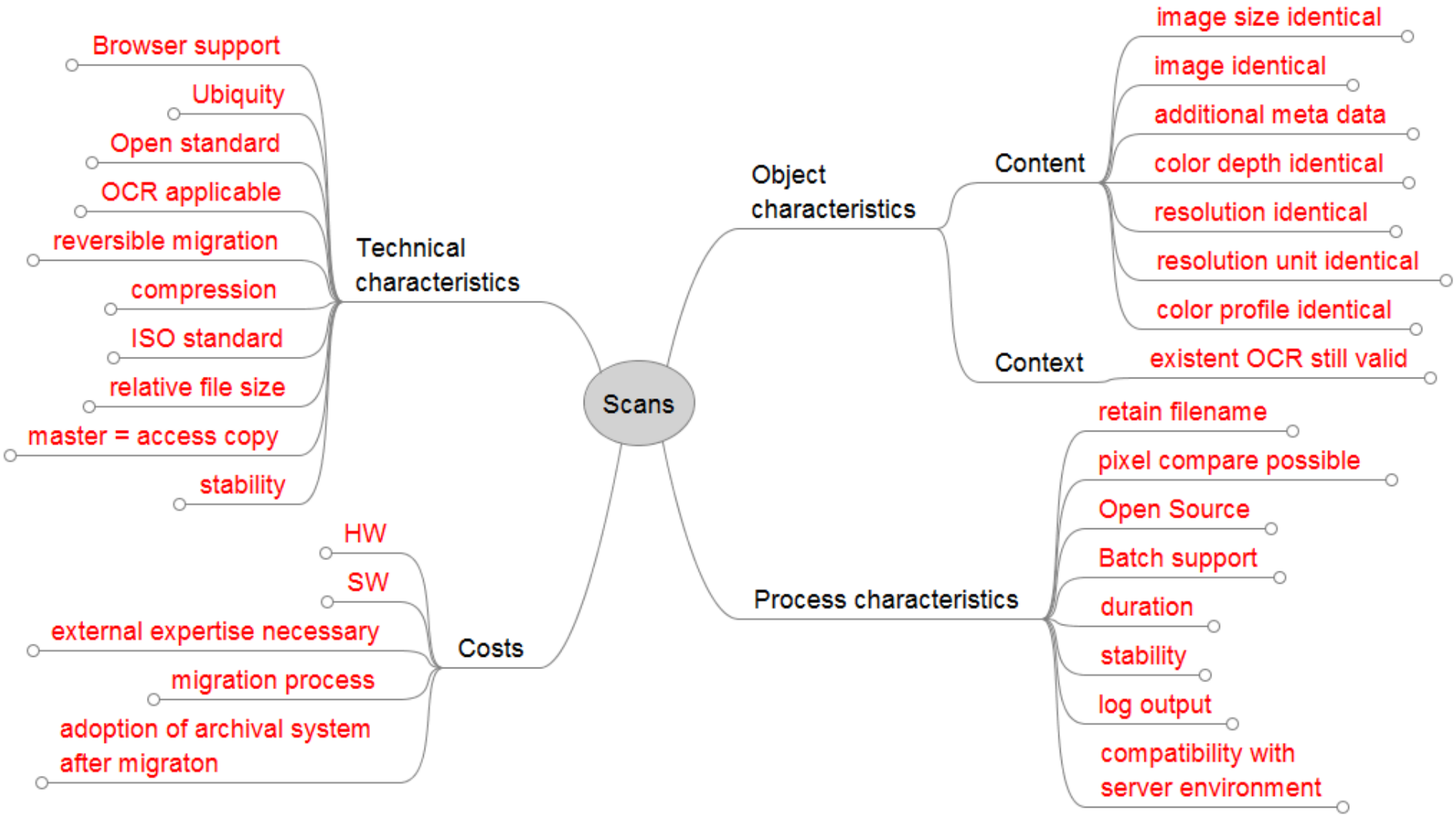
Software quality in preservation scenarios

- ISO 25010 SQUARE: Software quality model
  - *(Business factors not part of SQUARE)*
- Need to define indicators and metrics for specific domain



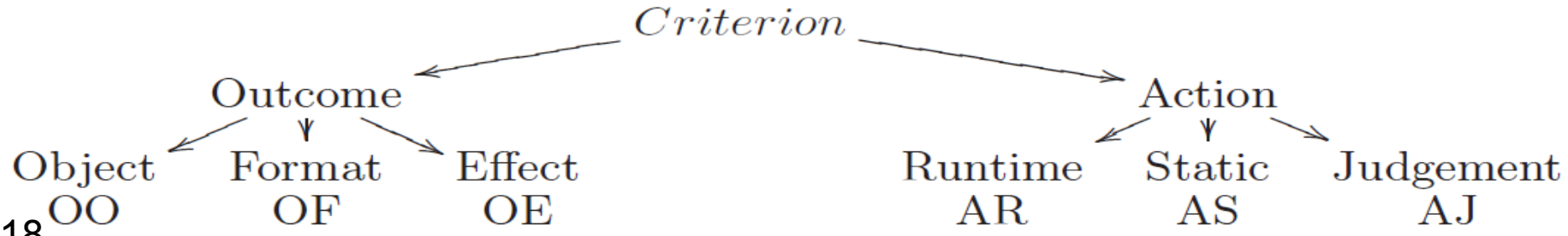

Action Properties
reliability
maturity
runtime stability
stability indicators
support availaible
under development
release cycle
active community
manufacturer status
availability
fault tolerance
error tolerance
error correction
recoverability
functional suitability
functional completeness
generic
number of input-formats supported
number of output-formats supported
macro support
input object validation
output file format verification
retain filename
activity traceability
activity logging format
activity logging amount
error reporting
understandability of reported errors
reported errors contain handling advice
number of objects supported
fraction of objects supported
keyboard support
mouse support
sound support
cd-rom support
content specific
functional correctness
refers to category outcome object
functional appropriateness
maintainability
modularity
modular design
plug-in system
reusability
analysability
modifiability
customization
testability
performance efficiency
time behaviour
time per MB
time per sample
throughput
resource utilization
memory per sample
memory per MB
resource utilization
capacity
batch processing maximum files
compatibility
co-existence
interoperability
interfaces
batch processing
directory watch
API
.Net API
Java API
Other API
portability
adaptability
installability
supported plattforms
ease of integration in current environment
replaceability
usability
appropriateness recognizability
learnability
documentation
availability
quality
operability
ease of use
ease of administration
configuration method
object configuration method
user error protection
user interface aesthetics
accessibility
business
licencing schema
costs
setup
execution
total

Scanned books requirements

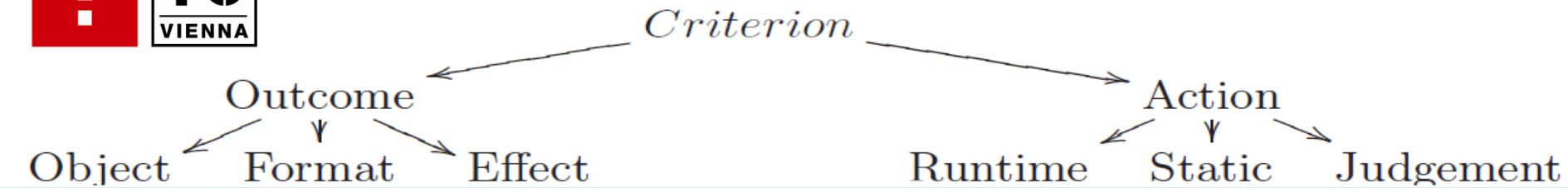


Decision criteria: What to measure?

- Each criterion concerns either the action or its outcome
- **Outcome**
  - **Object** (authenticity, editability, …)
  - **Format** (licensing, standardisation, complexity…)
  - **Effect** (Costs…)
- **Action**
  - **Runtime** properties (performance, stability, logging…)
  - **Static** (price, license…)
  - **Judgement** (configuration interface usability…)

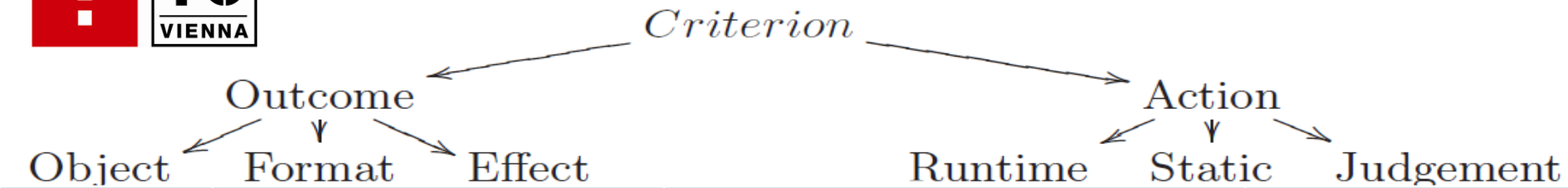


How to measure?



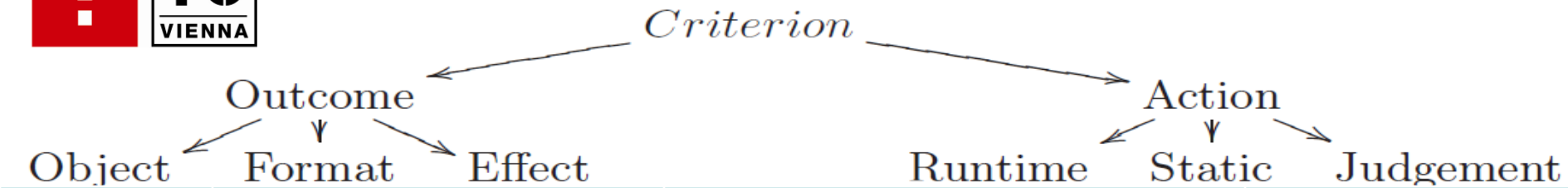
Category	Example	Data collection and measurement	Tools

How to measure?



Category	Example	Data collection and measurement	Tools
Outcome Object	**Image pixelwise identical Footnotes preserved**	**Measurements of output and input, comparison**	**FITS, JHove, image comparison...**

How to measure?



Category	Example	Data collection and measurement	Tools
Outcome Object	Image pixelwise identical Footnotes preserved	Measurements of output and input, comparison	FITS, JHove, ImageMagick...
Outcome Format	**Format is ISO standardised**	**Measurements of the output, Trusted external data sources**	**DROID, PRONOM, UDFR, P2**

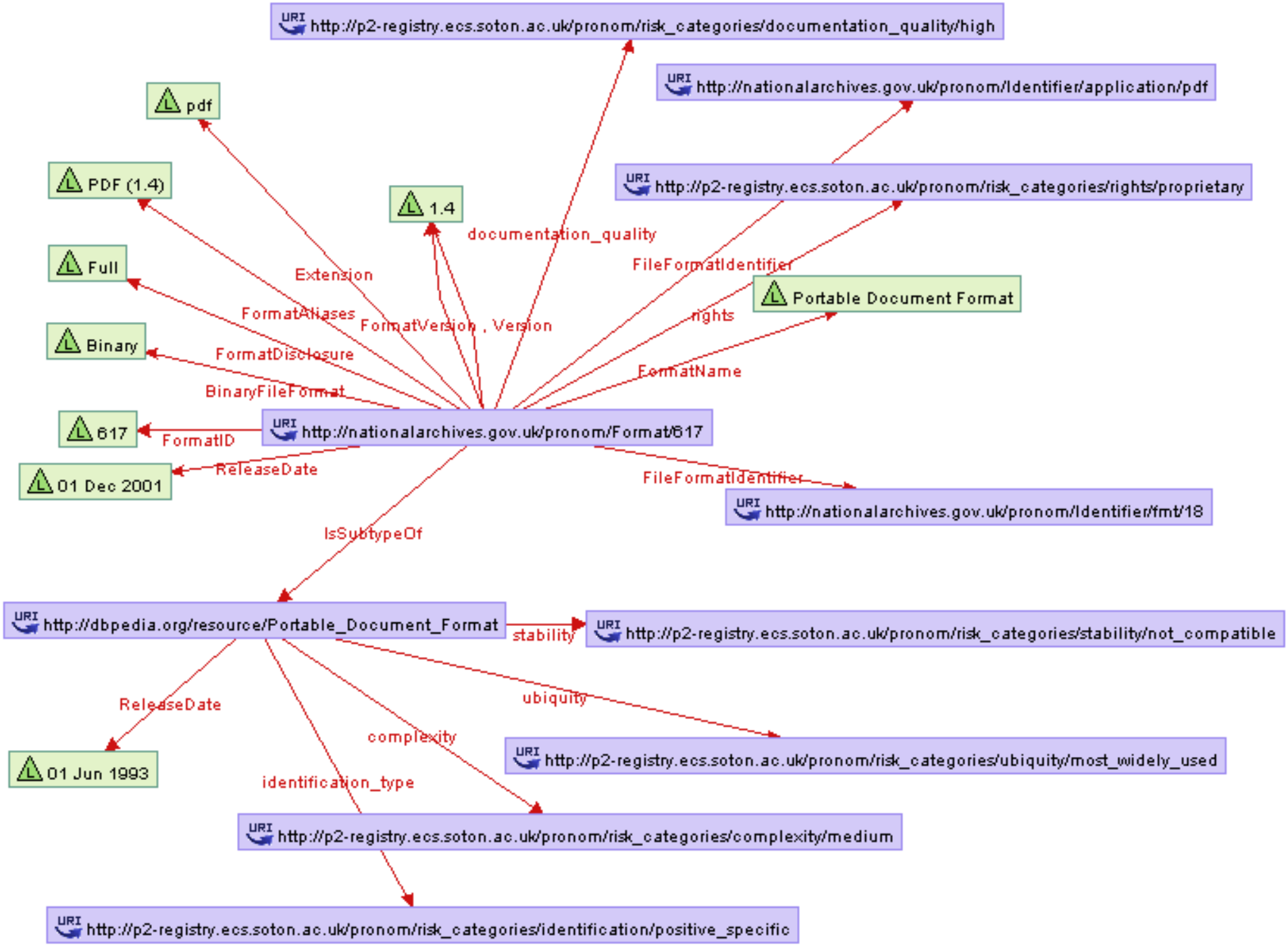
Some file format requirements

- Specifications available
  - Is an XML schema enough?
  - Syntacs **and semantics** needed
- Standardized (ISO, ANSI, ITEF, ...)
- Accepted and widely used (indicators?)
- Not covered by patent
- Free of compression
- Free of any cryptographical techniques

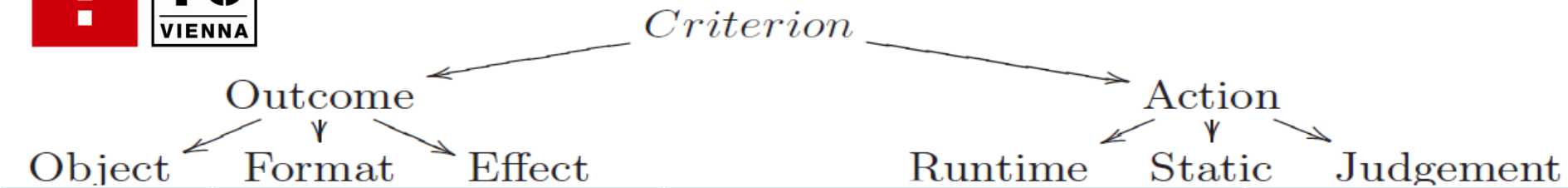
- Flexible and extensible?

Data sources

- PRONOM
  - Sparse data
- www.digitalpreservation.gov/formats
  - Incomplete
- Wikipedia
  - reliable?
- The web
  - unstructured
- P2: Combination of PRONOM with dbpedia
  - Linked Data
  - ~45.000 statements
  - Still far from complete

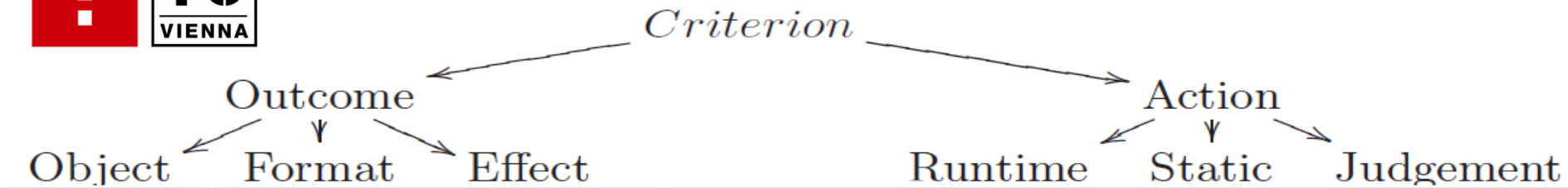

http://p2-registry.ecs.soton.ac.uk/pronom/risk_categories/documentation_quality/high
http://nationalarchives.gov.uk/pronom/Identifier/application/pdf
pdf
PDF (1.4)
1.4
http://p2-registry.ecs.soton.ac.uk/pronom/risk_categories/rights/proprietary
documentation_quality
Full
Extension
FileFormatIdentifier
Portable Document Format
FormatAliases
FormatVersion , Version
rights
Binary
FormatDisclosure
FormatName
BinaryFileFormat
617
FormatID
http://nationalarchives.gov.uk/pronom/Format/617
ReleaseDate
01 Dec 2001
FileFormatIdentifier
http://nationalarchives.gov.uk/pronom/Identifier/fmt/18
IsSubtypeOf
http://dbpedia.org/resource/Portable_Document_Format
stability
http://p2-registry.ecs.soton.ac.uk/pronom/risk_categories/stability/not_compatible
ReleaseDate
ubiquity
01 Jun 1993
complexity
http://p2-registry.ecs.soton.ac.uk/pronom/risk_categories/ubiquity/most_widely_used
identification_type
http://p2-registry.ecs.soton.ac.uk/pronom/risk_categories/complexity/medium
http://p2-registry.ecs.soton.ac.uk/pronom/risk_categories/identification/positive_specific

How to measure?



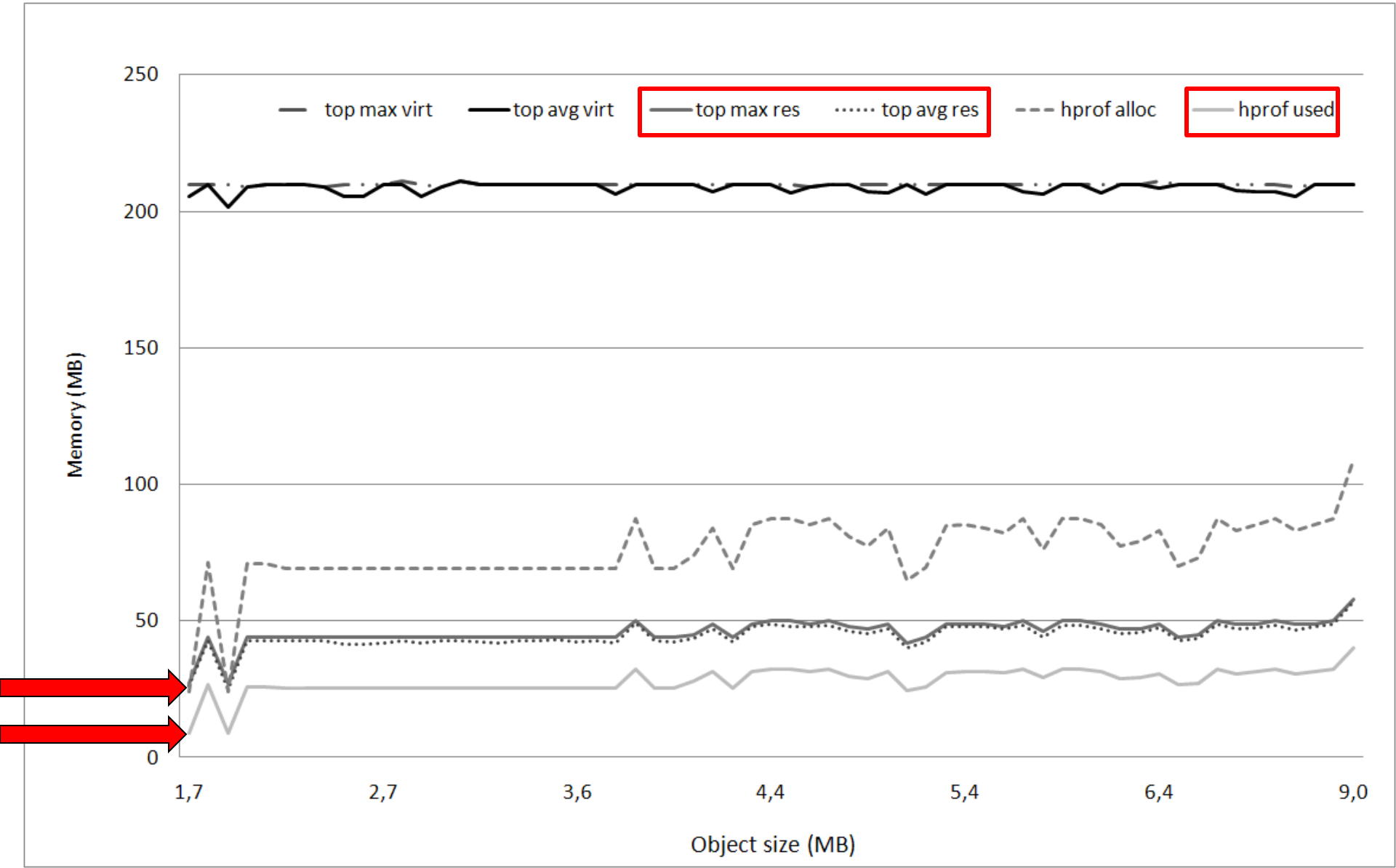
Category	Example	Data collection and measurement	Tools
Outcome Object	Image pixelwise identical Footnotes preserved	Measurements of output and input, comparison	FITS, JHove, ImageMagick...
Outcome Format	Format is ISO standardised	Measurements of the output, Trusted external data sources	DROID, PRONOM, UDFR, P2
Outcome effect	**Annual bitstream preservation costs (€)**	**Measurements of the output, external data sources, models (LIFE)...**	**LIFE model**

How to measure?

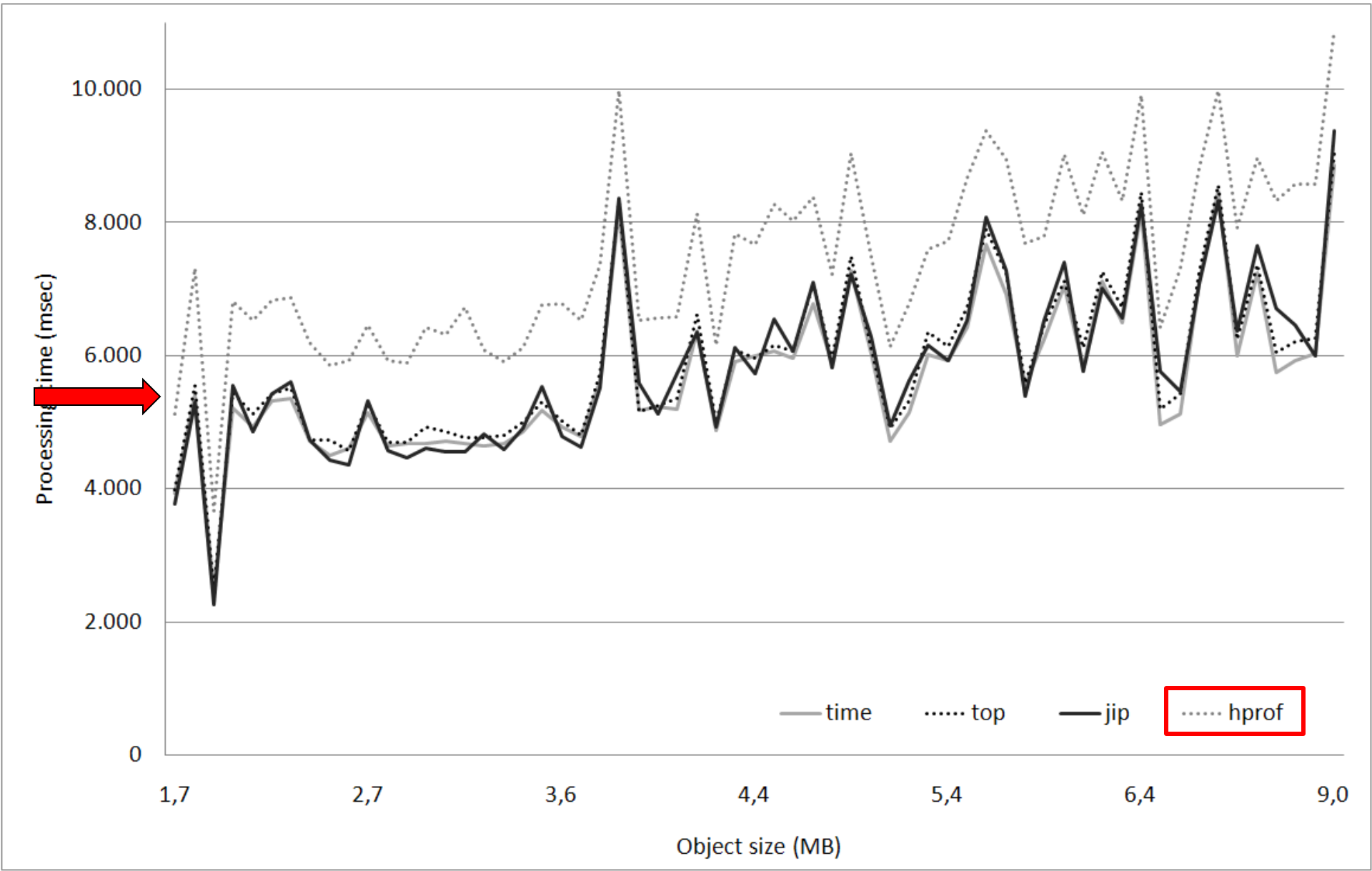


Category	Example	Data collection and measurement	Tools
Outcome Object	Image pixelwise identical Footnotes preserved	Measurements of output and input, comparison	FITS, JHove, ImageMagick...
Outcome Format	Format is ISO standardised	Measurements of the output, Trusted external data sources	DROID, PRONOM, UDFR, P2
Outcome effect	Annual bitstream preservation costs (€)	Measurements of the output, external data sources, models (LIFE)...	LIFE model
Action runtime	**Throughput (MB per millisecond), Memory usage**	**Measurements taken in controlled experimentation**	**MiniMEE**

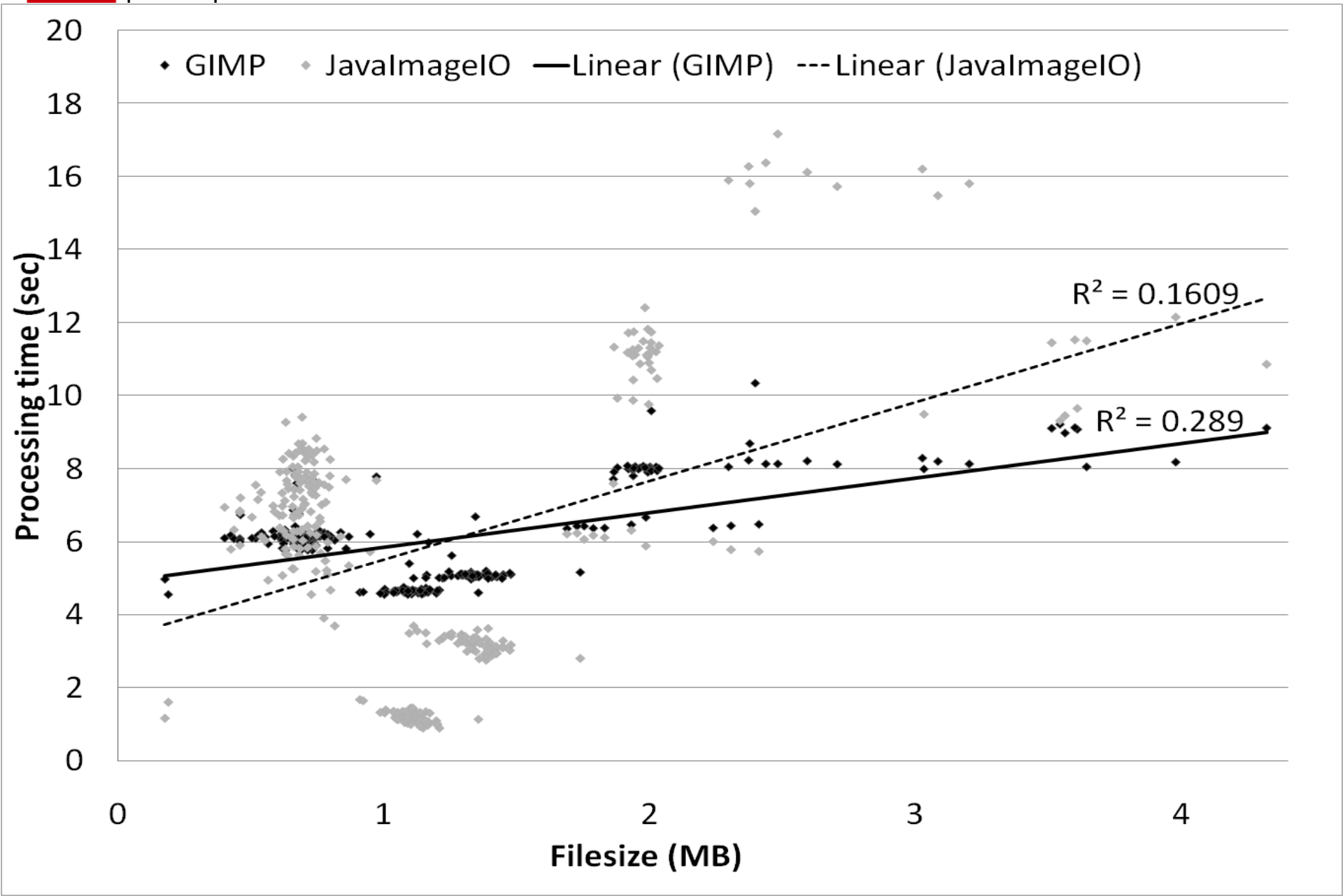
Profiling memory usage of Java tools



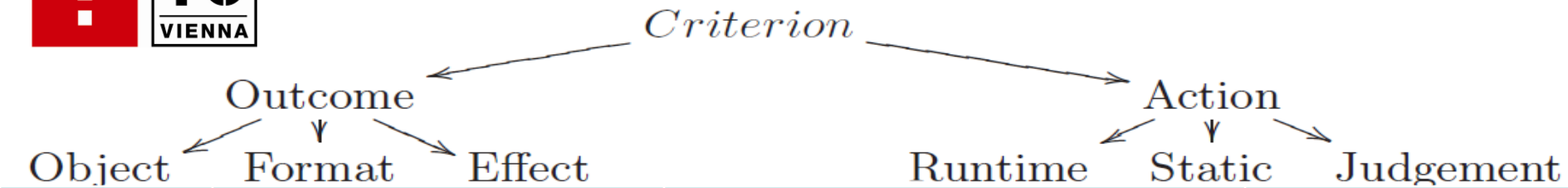
Profiling timing of Java tools



Comparing tool performance

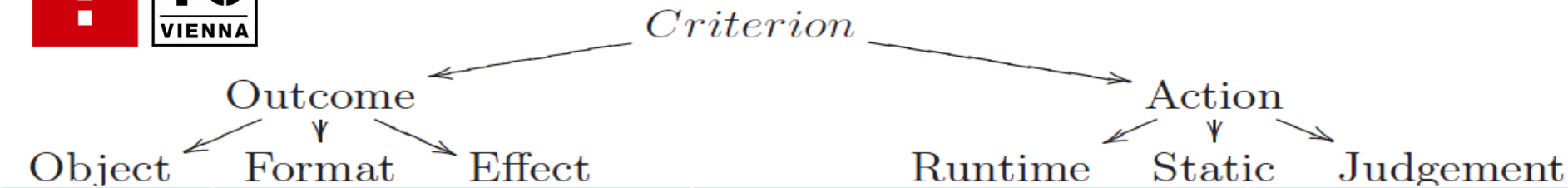


How to measure?



Category	Example	Data collection and measurement	Tools
Outcome Object	Image pixelwise identical Footnotes preserved	Measurements of output and input, comparison	FITS, JHove, ImageMagick...
Outcome Format	Format is ISO standardised	Measurements of the output, Trusted external data sources	DROID, PRONOM, UDFR, P2
Outcome effect	Annual bitstream preservation costs (€)	Measurements of the output, external data sources, models (LIFE)...	LIFE model
Action runtime	Throughput (MB per millisecond), Memory usage	Measurements taken in controlled experimentation	MiniMEE
Action static	**License costs per CPU (€), Open Source License**	**Trusted external data sources, manual evaluation, sharing**	**UDFR, Pronom, P2, manual**

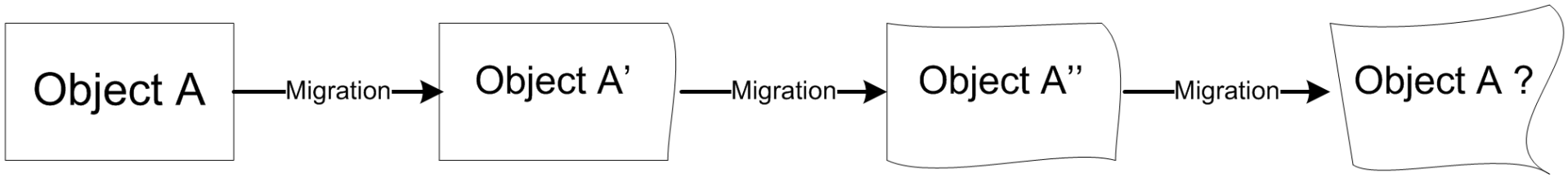
How to measure?



Category	Example	Data collection and measurement	Tools
Outcome Object	Image pixelwise identical Footnotes preserved	Measurements of output and input, comparison	FITS, JHove, ImageMagick...
Outcome Format	Format is ISO standardised	Measurements of the output, Trusted external data sources	DROID, PRONOM, LoC format site, UDFR, P2
Outcome effect	Annual bitstream preservation costs (€)	Measurements of the output, external data sources, models (LIFE)...	LIFE model
Action runtime	Throughput (MB per millisecond), Memory usage	Measurements taken in controlled experimentation	MiniMEE
Action static	License costs per CPU (€), Open Source License	Trusted external data sources, manual evaluation, sharing	UDFR, P2, manual
Action judgement	**Technical interoperability, configuration flexibility**	**Manual judgement, sharing**	

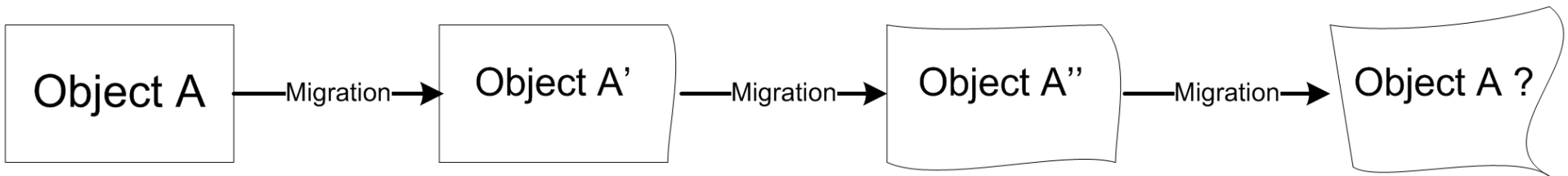
Core requirement: Keep object intact

- Essential object characteristics
  - Content
  - Appearance
  - Structure
  - Behaviour
  - Context

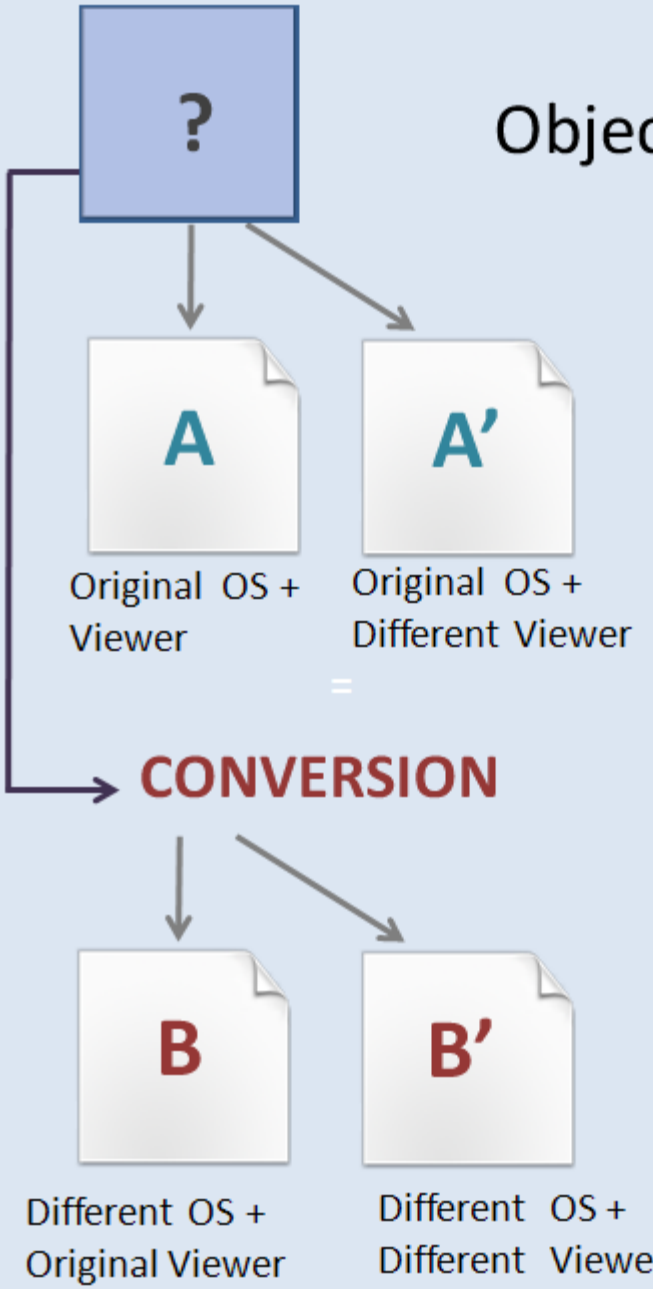


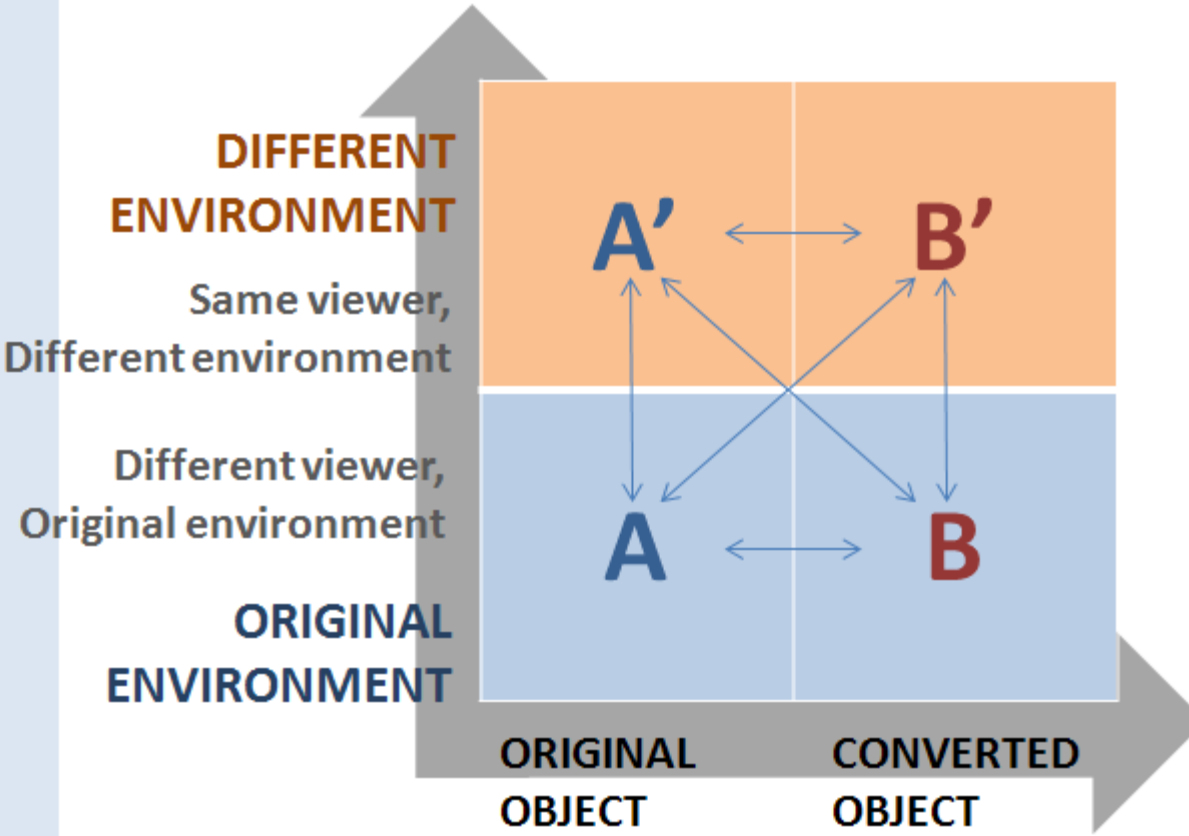
Validating a migrated image

- We run a TIFF file against a JPEG 2000 conversion tool and check:
- ✓ Yes, it's in JPEG 2000 format
- ✓ Yes, it's well-formed
- ✓ Yes, it's valid
- ✓ Yes, it still has the same dimensions
- …. But is it still the same image?



Objects, environments and dependencies

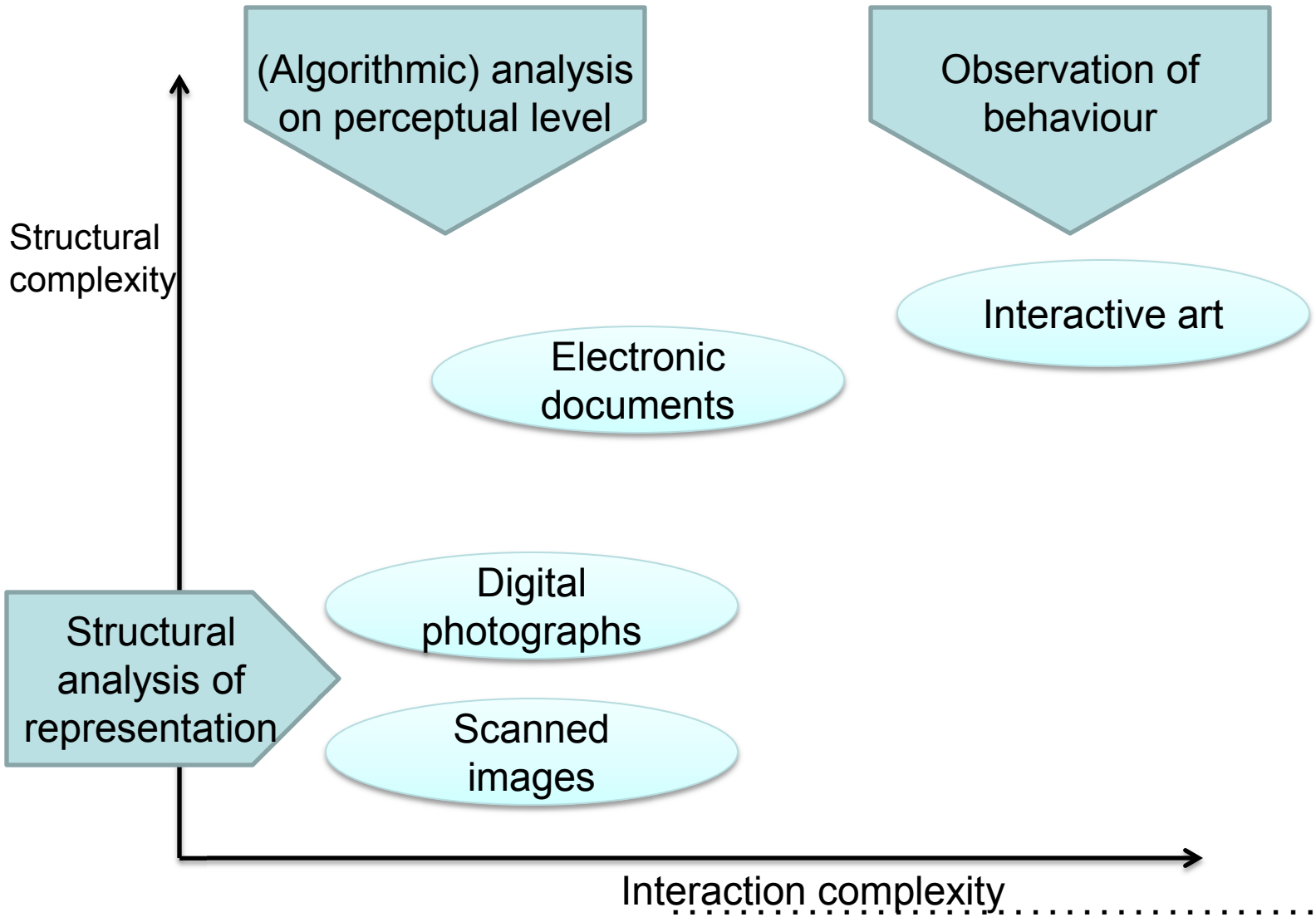




...networks of objects

(Diagram by Natasa Milic-Frayling, MSRC)

Approaches to Quality Assurance



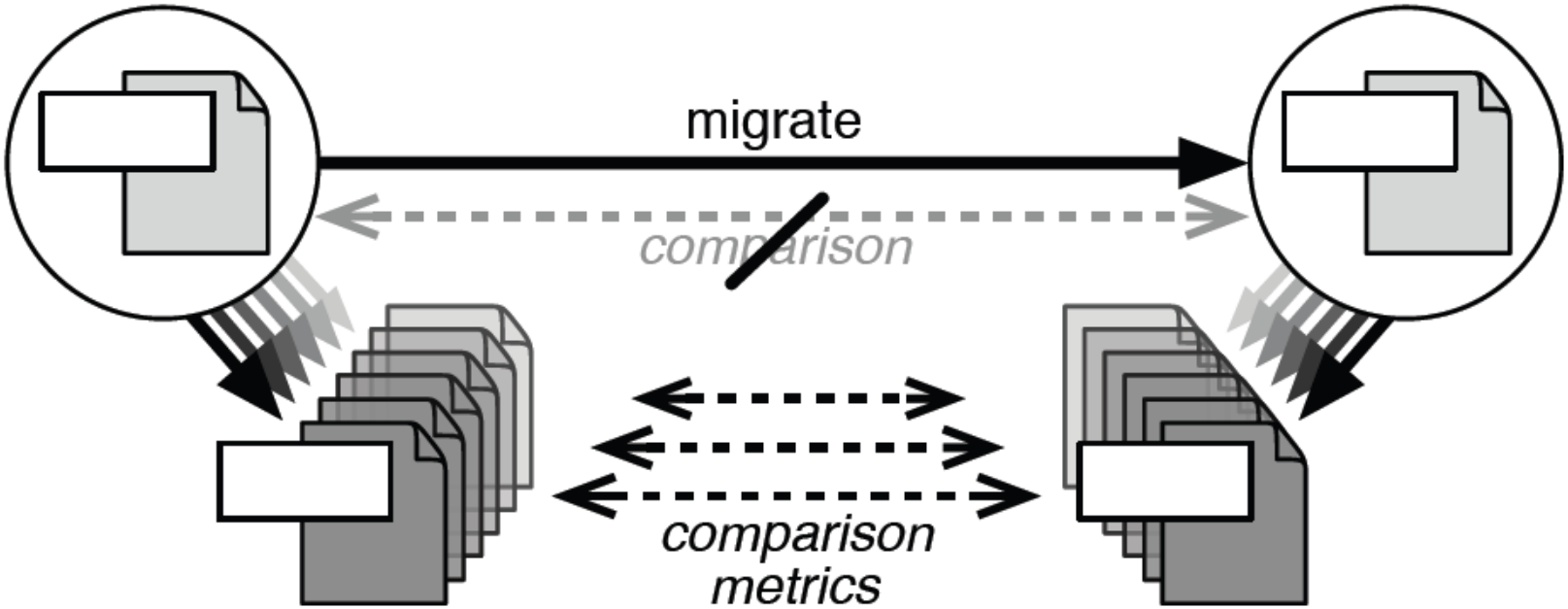
Validating a migrated image

- Dimensions, metadata.... easy: extract and compare
- Content... Not always easy
- ImageMagick *compare:* good for simple cases

Abbr.	Metric	Description
AE	Absolute Error	The number of different pixels (0 means identical images). This value can be thresholded to only count pixels that have a difference larger then a specified threshold.
PAE	Peak Absolute Error	The highest difference of any single pixel.
PSNR	Peak Signal to Noise Ratio	The ratio of mean square difference to the maximum mean square that can exist between any two images, expressed as a decibel value. The higher the PSNR, the closer the images are, with a maximum difference occurring at 1.
MAE	Mean Absolute Error	Average over all pixels
MSE	Mean Squared Error	Averaged squared error distance
RMSE	Root mean squared error	Identical to *sqrt(MSE)*.

Evaluating conversion processes

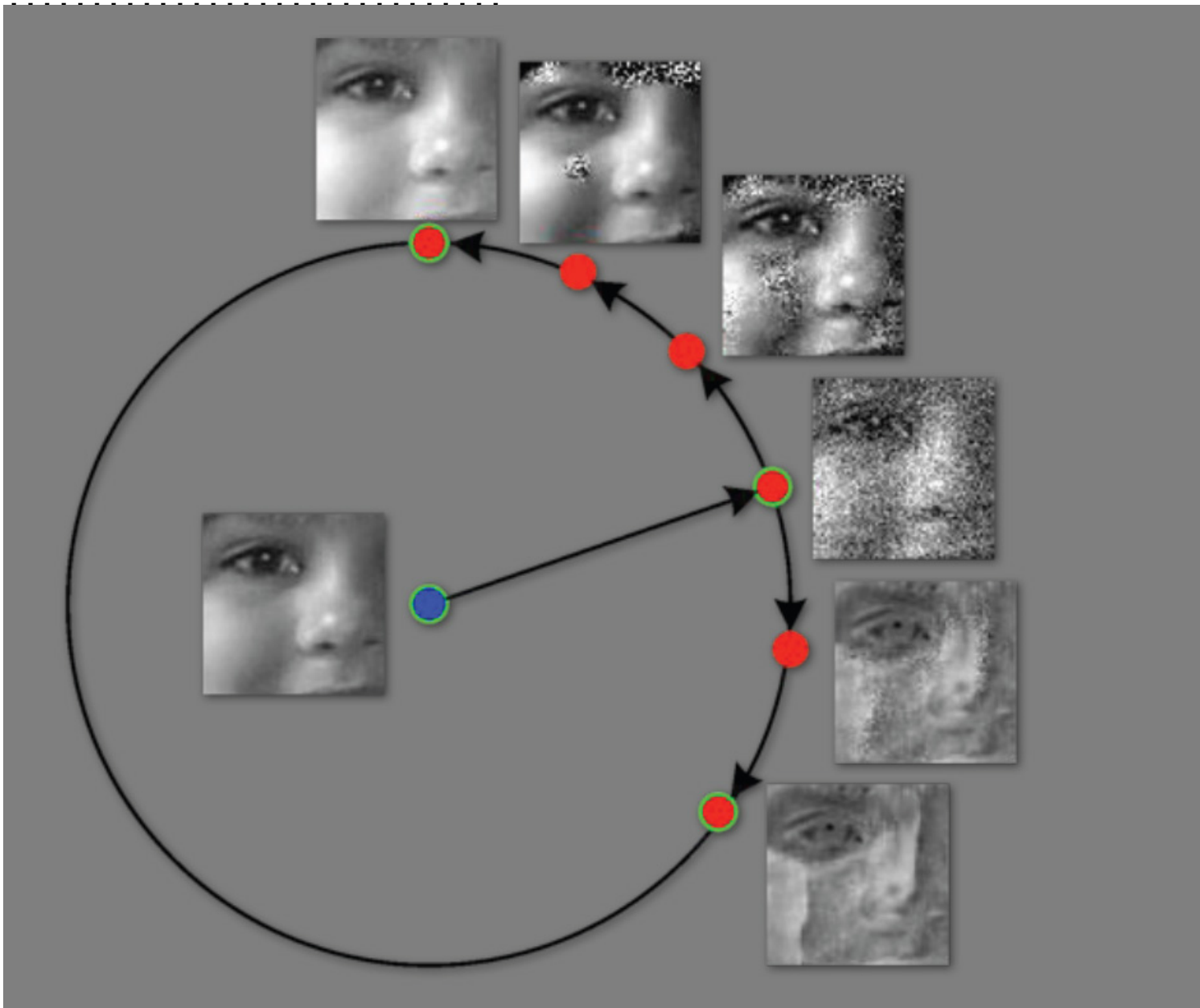
- Approaches to automated Quality Assurance
  - Characterization: Static analysis of representations
    - Semantics not fully contained in files
    - No homomorphic property mapping between (lots of) formats
    - Lack of ground truth
  - Perceptual-level analysis of standardised renderings



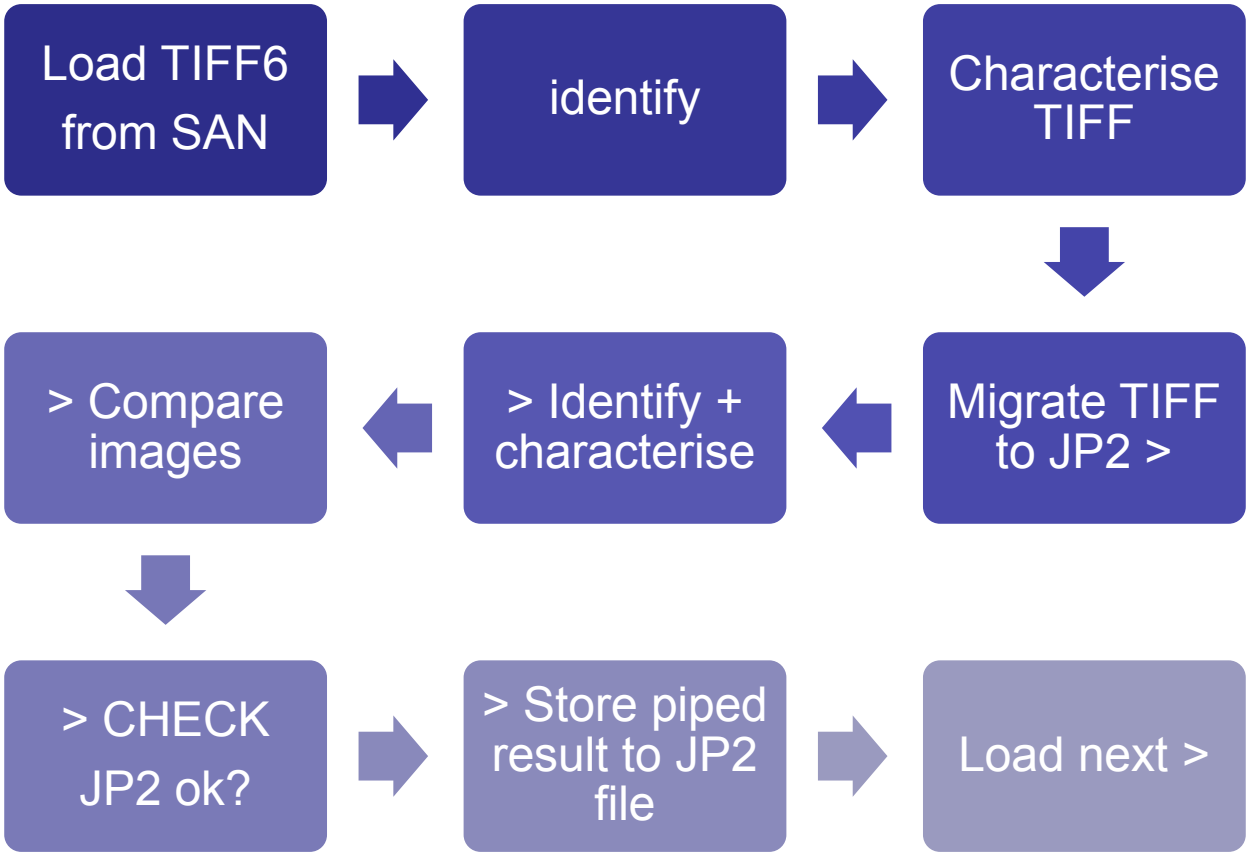
Distance metrics: How meaningful?

AE
PAE
RMSE
...
SSIM

Anything but “0” is a problematic result



A real migration workflow... What's wrong?



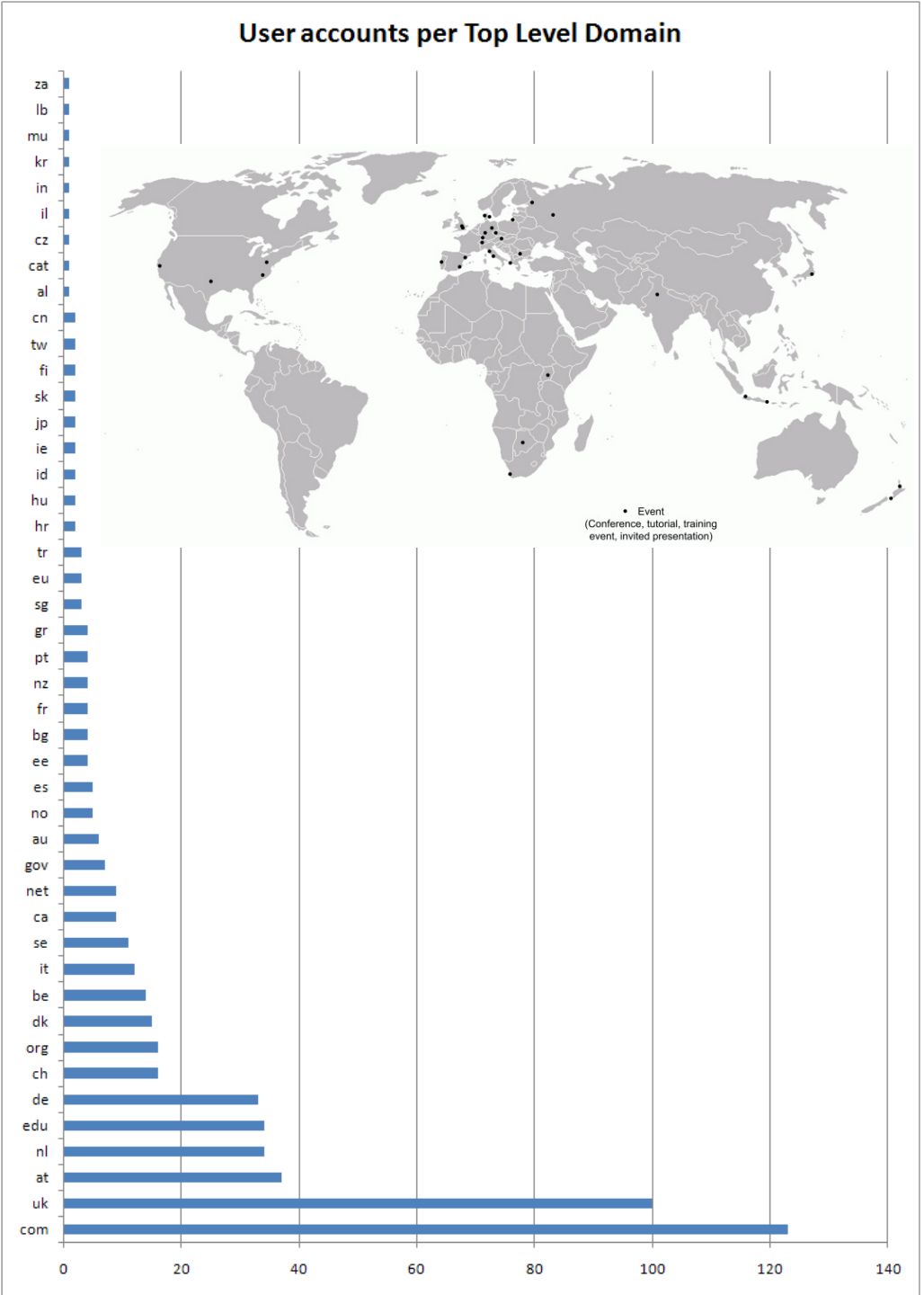
Criteria and solutions

- Problem space vs. solution space
- Criteria for objects, representation instances and formats
  - **Significant properties**: The content of the objects that has to be preserved in order to preserve authenticity.
    See "*Data without meaning: Establishing the Significant Properties of Digital Research". Gareth Knight and Maureen Pennock, IJDC 4(1) 2009.* *http://www.ijdc.net/index.php/ijdc/article/view/110/87*
  - **Representation instance properties***:* Aspects of the representation such as encoding, encryption, bytestream length, compression, validity, profile conformance. See http://publik.tuwien.ac.at/files/PubDat_203384.pdf
  - **Format properties**: Properties of the formats themselves, not of the objects (status of standardisation, support by viewing environments, license, tool support, features, versions,…) www.digitalpreservation.gov/formats

Conclusions: Preservation planning

- From strategy and policies to operations
- A simple, methodologically sound model to specify and document requirements
- Repeatable and documented evaluation for informed and accountable decisions
- Generic workflow that can be integrated in different institutional settings
- **Plato**:
  Tool support to perform solid, well-documented analysis
- Provides basic preservation plan

http://www.ifs.tuwien.ac.at/dp/plato



Plato

- De-facto standard preservation planning tool
- ~1000 user accounts
- Case studies
- Opportunities to leverage scale
- Analyse data set from diverse communities and build improvements

www.ifs.tuwien.ac.at/dp/plato

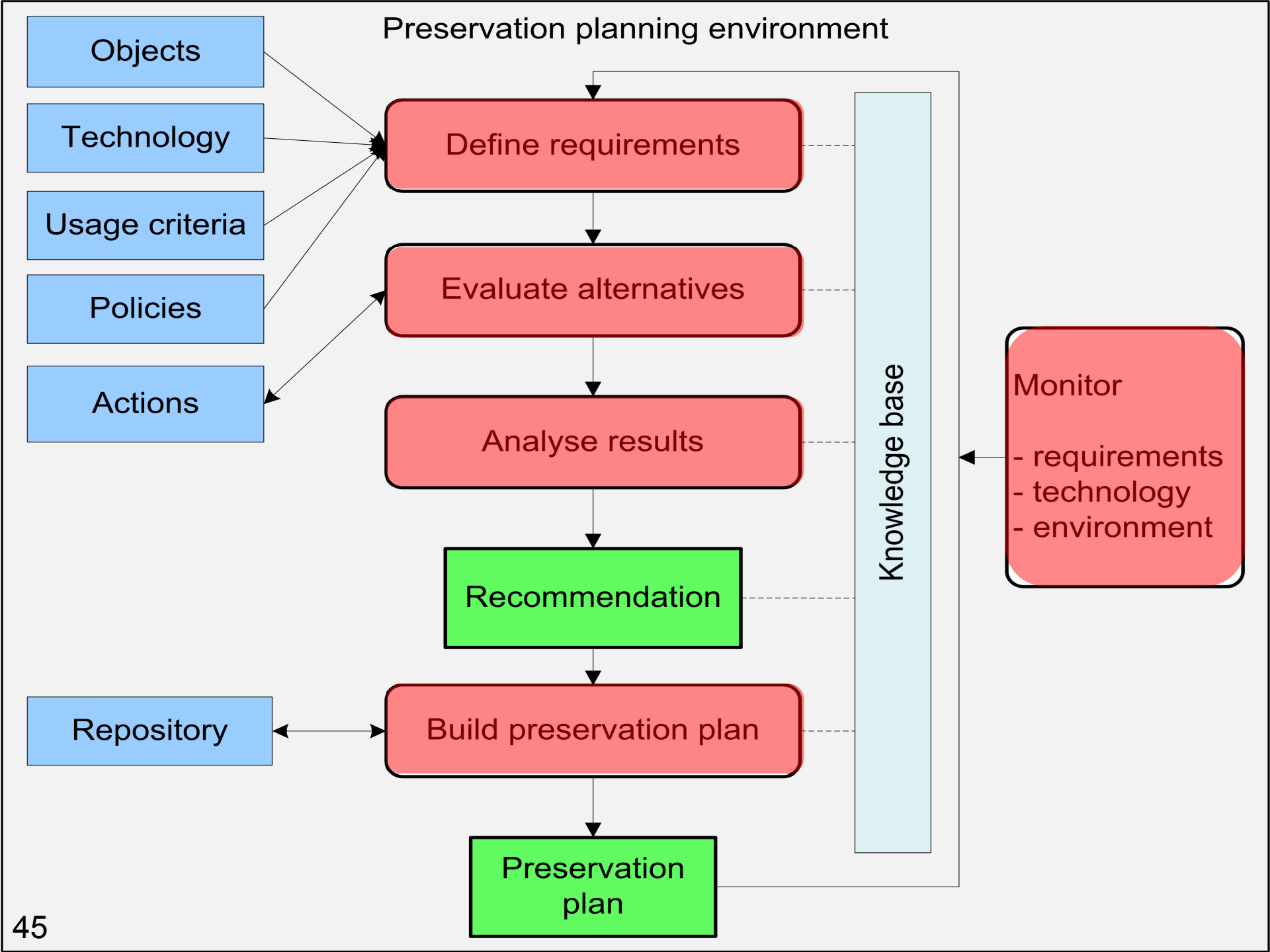
Planning Challenges

- Creating a plan is effort-intensive
- Sharing experience is difficult
- Monitoring changes is manual
- Integrating context, strategies and operations is difficult

Planning Challenges

- Creating a plan is effort-intensive
  - Increase efficiency of planning
- Sharing experience is difficult
  - Increase standardisation and reusability
- Monitoring changes is manual
  - Introduce automation
- Integrating context, strategies and operations is difficult
  - Systematic governance structures
  - Manage policies
  - Integrate systems

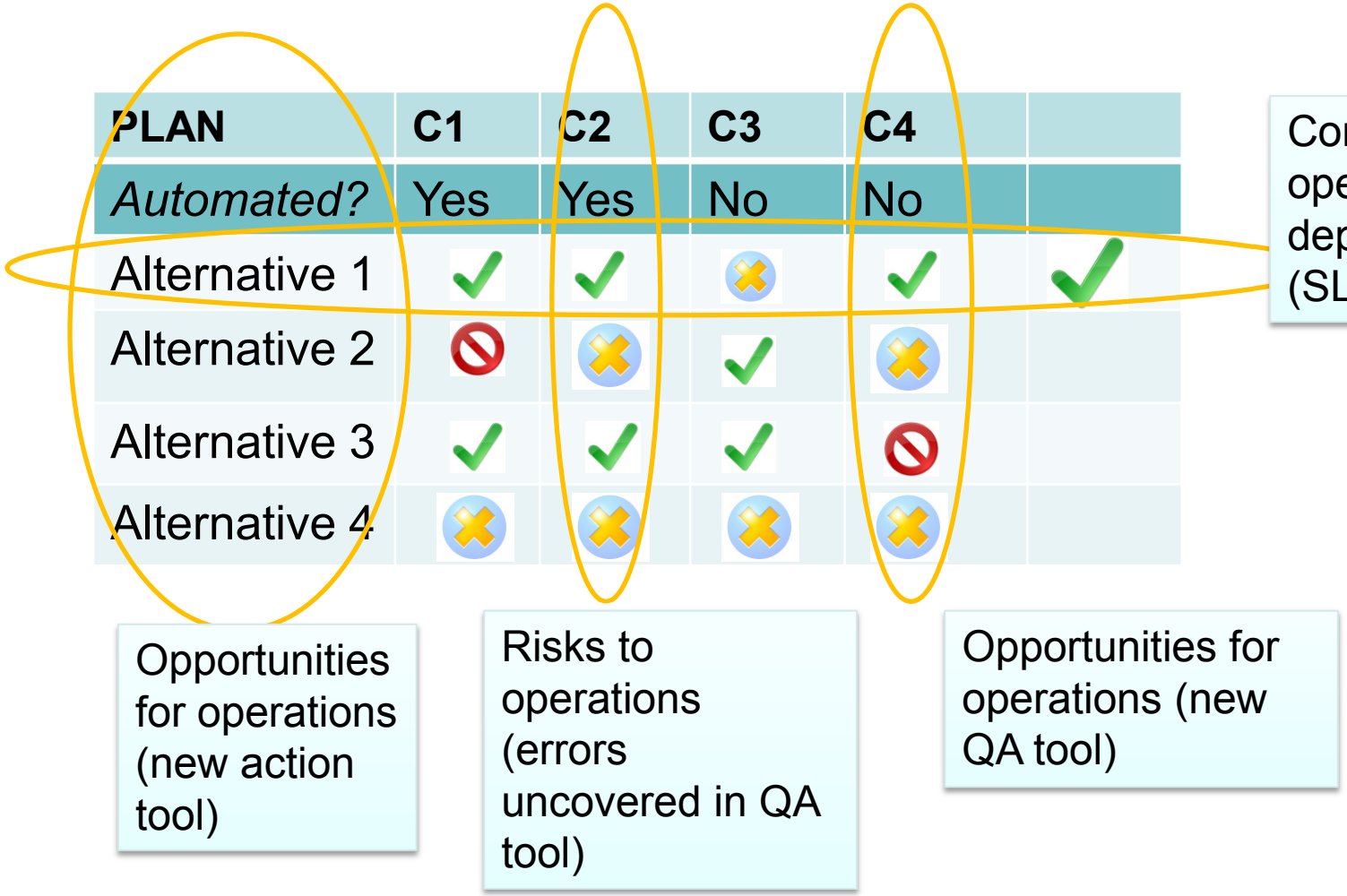
SCAPE
Scalable Preservation Environments


Preservation planning environment
Objects
Technology
Usage criteria
Policies
Actions
Define requirements
Evaluate alternatives
Analyse results
Recommendation
Build preservation plan
Preservation plan
Repository
Knowledge base
Monitor
- requirements
- technology
- environment

An Automated Monitoring service

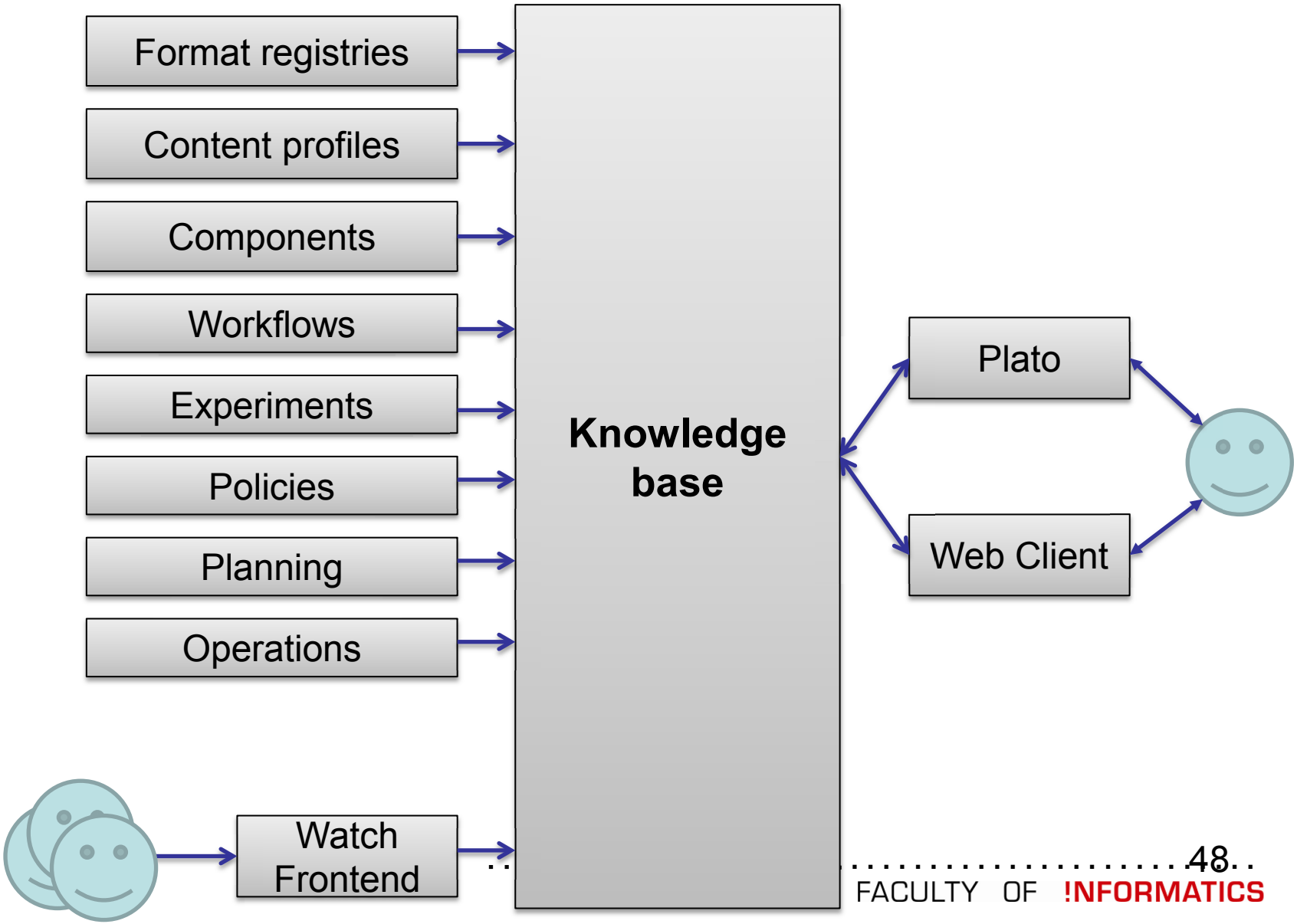
- Knowledge base
  - Entities and their properties
  - Measures of properties over time
  - Triggers define conditions and events
- Flexible and extensible
  - A well-defined, flexible data model
  - Adaptors for different information sources
- Monitoring Capabilities
  - Internal Monitoring
  - External Monitoring
  - Monitor compliance, risks and opportunities

Compliance, risk and opportunities

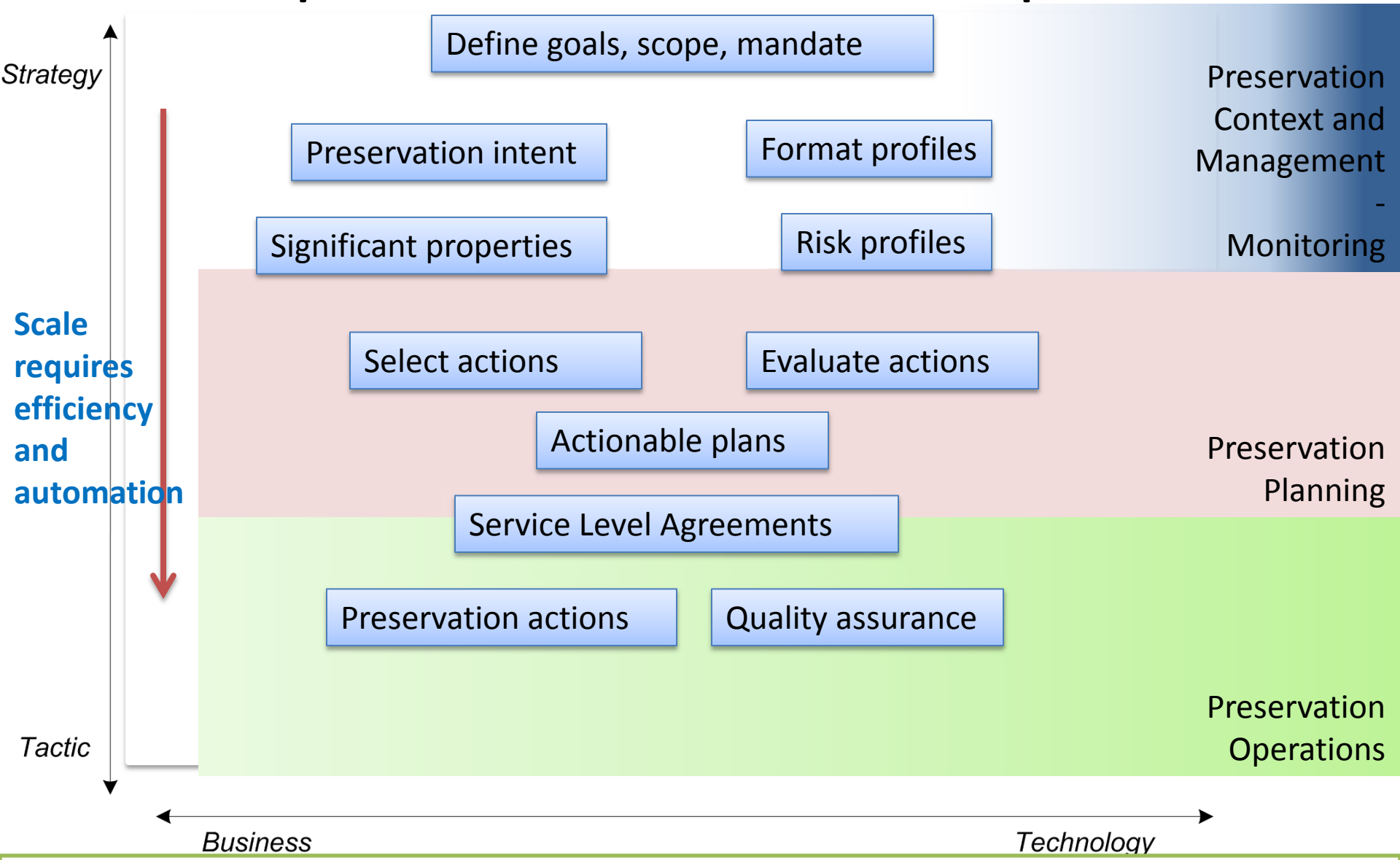


- Planning needs to generate Service Level Agreements (SLAs) and monitoring conditions

Information sources to be monitored



A preservation decision space

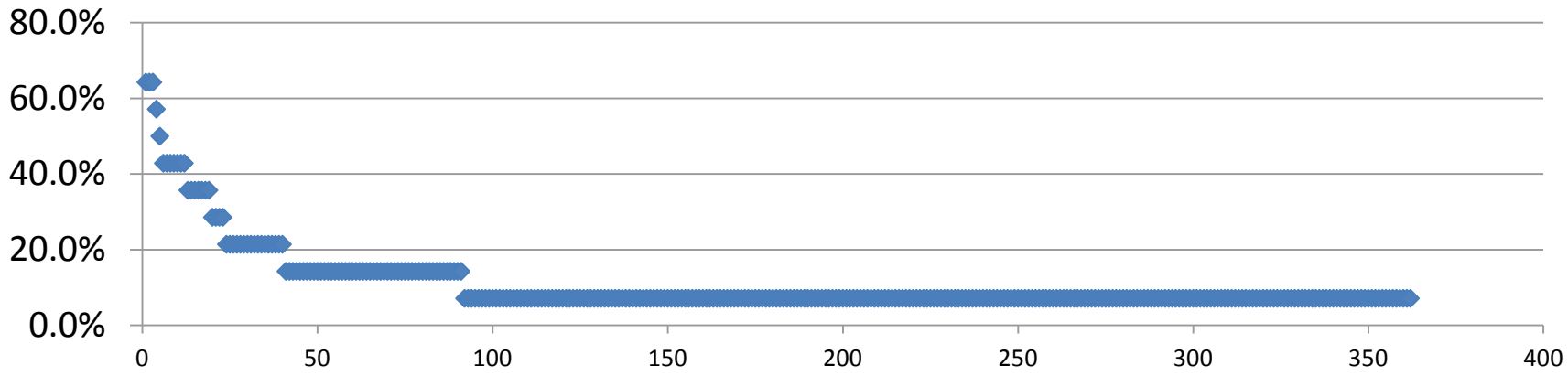


Goal: Efficiency and effectiveness

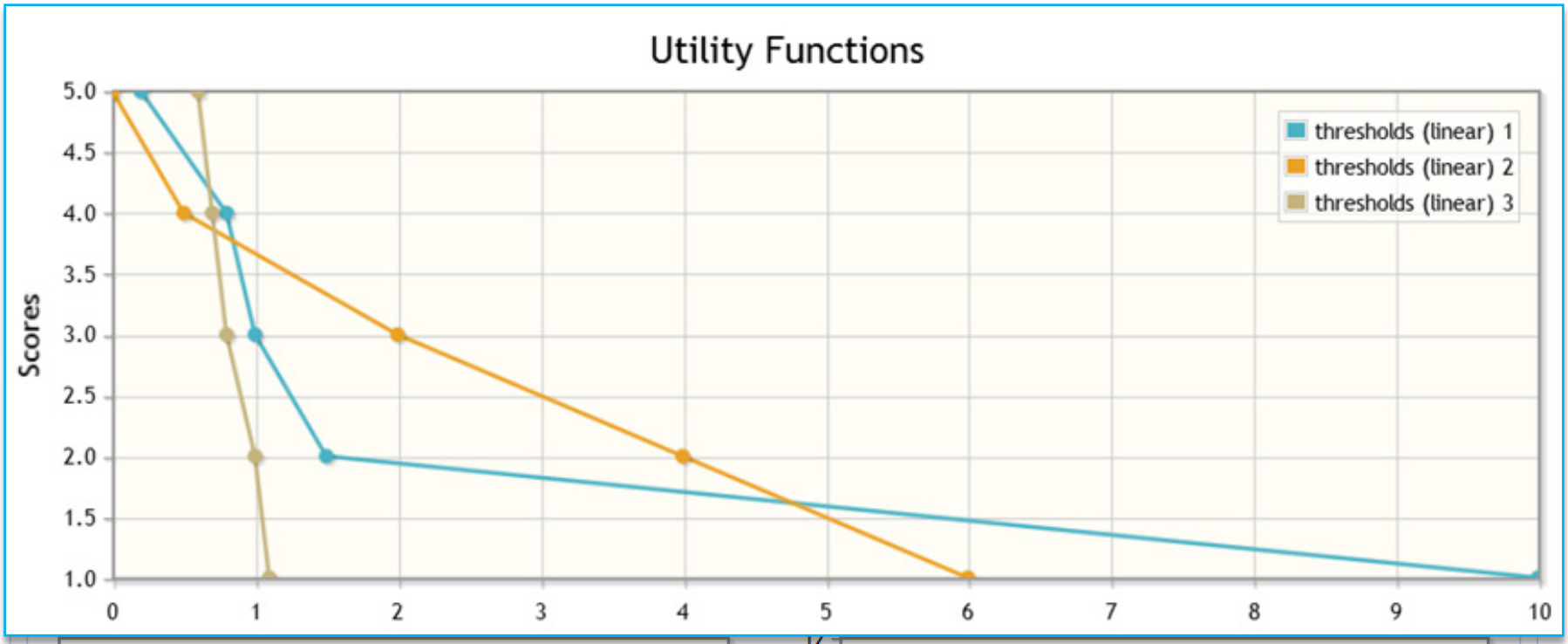
- Challenges
  - Effort for specifying criteria
  - Effort to perform evaluation
- We need effective and efficient near-optimal decisions
- Idea: Leverage scale
  - Standardize criteria
  - Develop quantitative metrics
  - Cross-reference and analyse
  - Provide visual insight
- Opportunity: Standardize, cross-reference, analyze

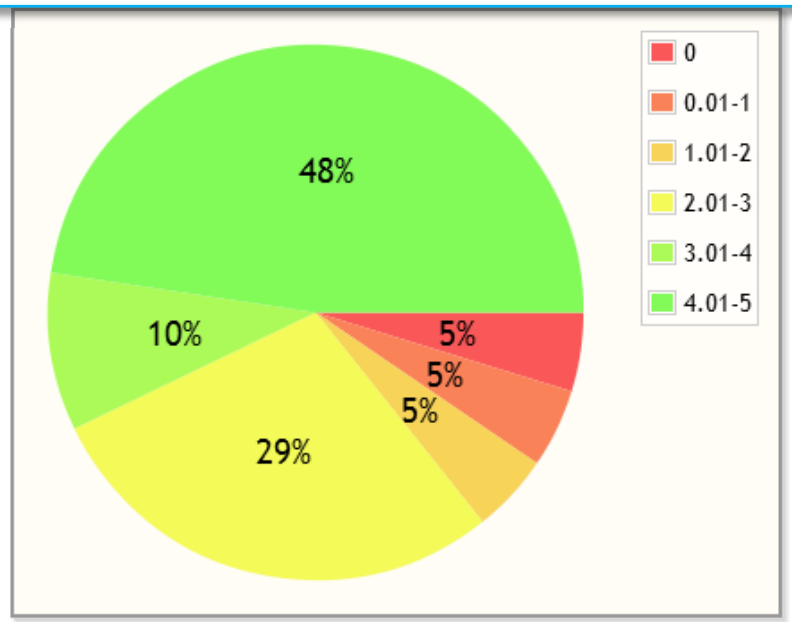
Collect

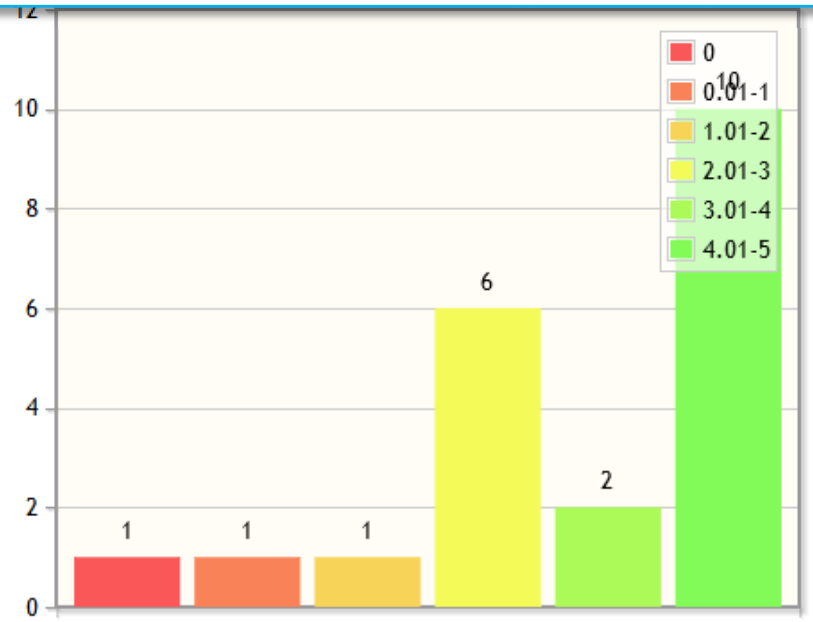
- 14 decision cases: Format conversion components
- 631 decision criteria (textually described)



Analyse criteria



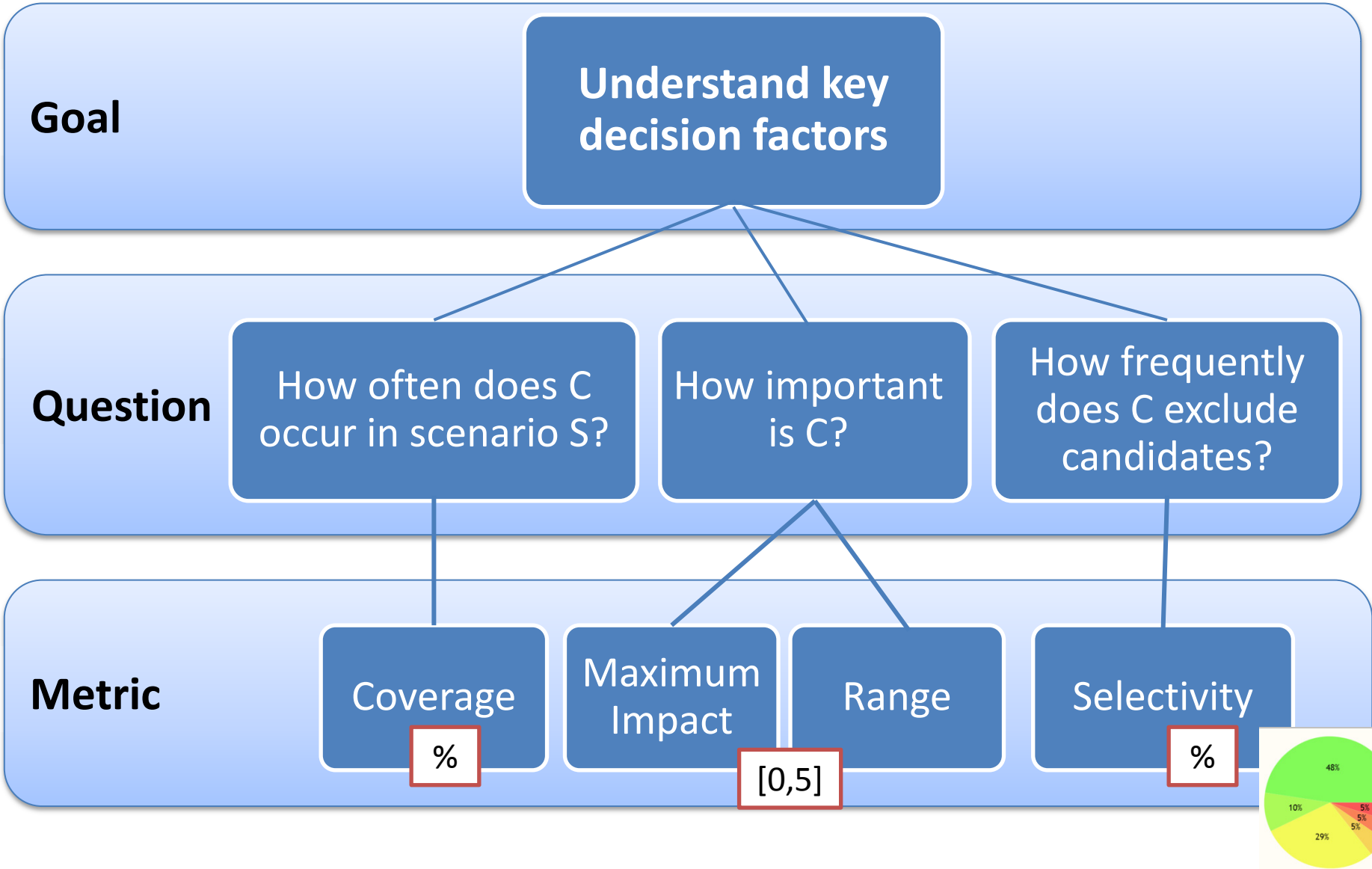




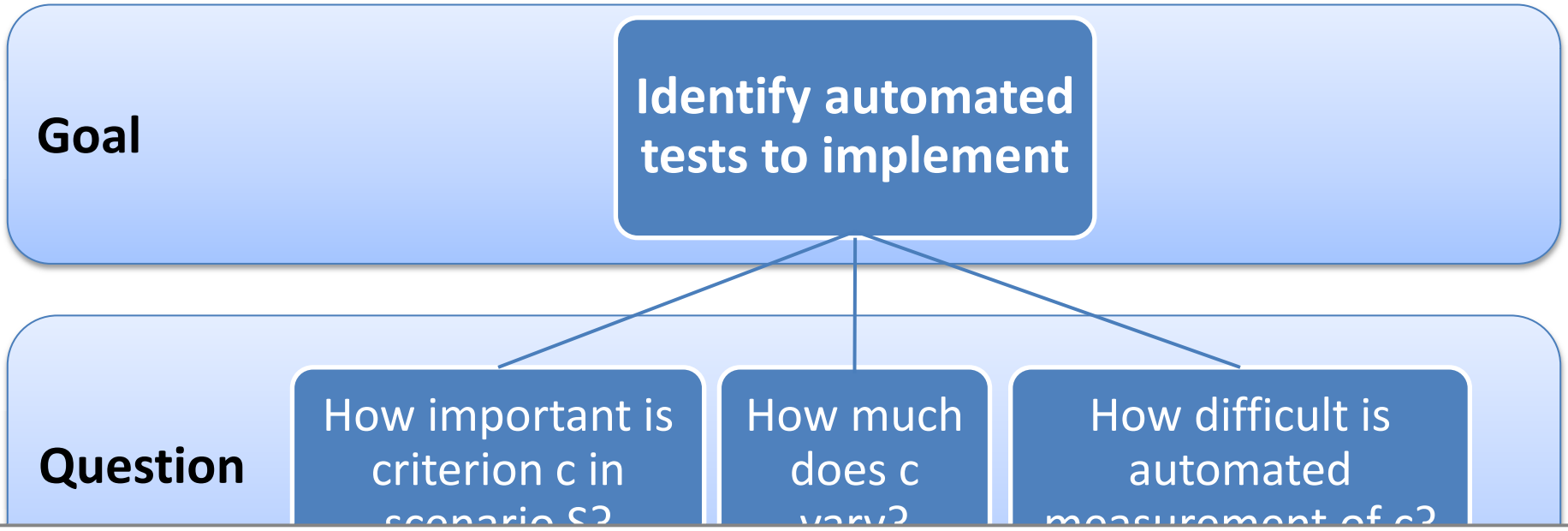
Analyze: Impact Factors

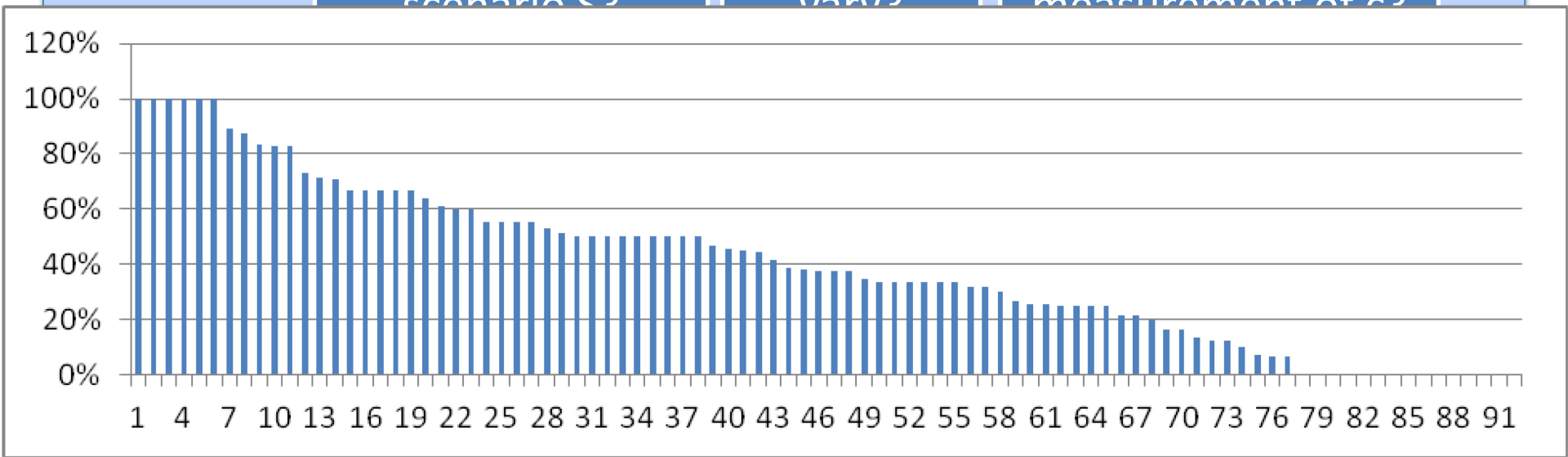
- Factors need to reflect
  - Usage frequency
  - Average weight
  - Sensitivity: How much do the utility scores change when the facts change?
    - Objective evidence
    - Utility function
  - Utility of 0 filters unacceptable candidates
- Using Goal-Question-Metric for selecting factors

Analyse a criterion (set) C



Automation for controlled experiments





Interpret

- Sets of criteria ranked by coverage

Criteria Sets Summary

Name	Size	Coverage	Max Impact	Range	Selectivity
SQ Functional Correctness (outcome object)	346	100%	3.457	0.95	13.79%
SQ Functional Correctness - TIP	195	100%	3.457	0.786	10.34%
Business	18	85.71%	0.33	0.155	2%
SQ Functional Completeness	44	78.57%	1.022	0.133	4.08%
SQ Portability	5	78.57%	0.5	0.054	0%
Format	31	71.43%	0.822	0.198	9.3%
SQ Performance Efficiency	7	64.29%	0.5	0.103	2.7%
SQ Resource Utilization	3	64.29%	0.5	0.103	2.7%
SQ Usability	6	64.29%	0.16	0.035	0%
SQ Functional Correctness - RIP	15	57.14%	0.156	0.03	2.7%
SQ Compatibility	5	57.14%	0.085	0.01	0%
SQ Time Behaviour	3	57.14%	0.5	0.094	3.33%
SQ Reliability	8	42.86%	0.154	0.016	0%
SQ Functional Correctness - Image Similarity	12	35.71%	0.401	0.056	8.33%
SQ Functional Correctness - IP	136	35.71%	0.625	0.135	12.5%
SQ Maintainability	3	28.57%	0.08	0.007	0%
Outcome Effects	4	21.43%	1.54	0.137	26.67%
SQ Capacity	1	0%	0	0	0%

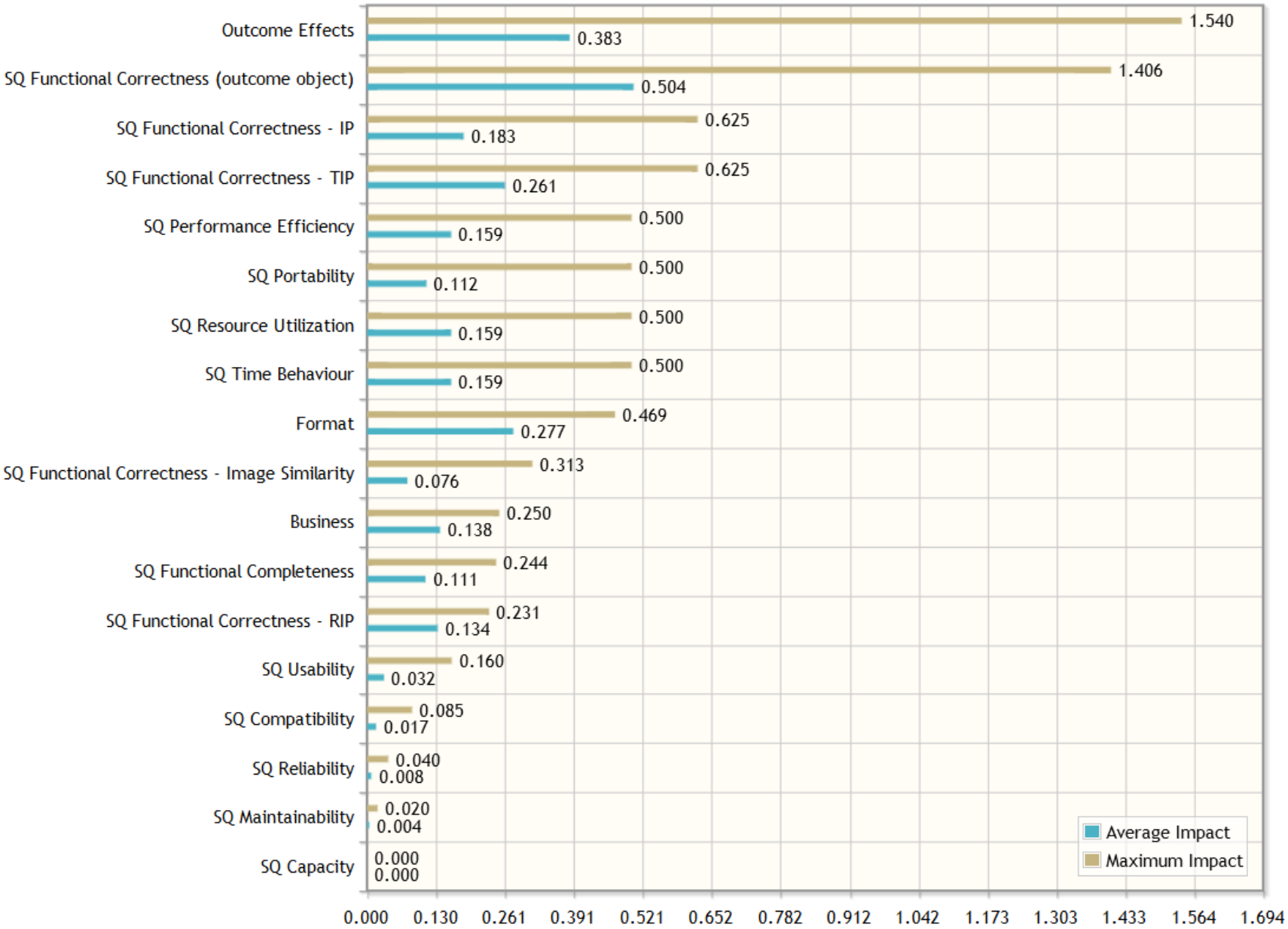
Interpret

- Sets of criteria ranked by selectivity

Criteria Sets Summary

Name	Size	Coverage	Max Impact	Range	Selectivity
Outcome Effects	4	21.43%	1.54	0.137	26.67%
SQ Functional Correctness (outcome object)	346	100%	3.457	0.95	13.79%
SQ Functional Correctness - IP	136	35.71%	0.625	0.135	12.5%
SQ Functional Correctness - TIP	195	100%	3.457	0.786	10.34%
Format	31	71.43%	0.822	0.198	9.3%
SQ Functional Correctness - Image Similarity	12	35.71%	0.401	0.056	8.33%
SQ Functional Completeness	44	78.57%	1.022	0.133	4.08%
SQ Time Behaviour	3	57.14%	0.5	0.094	3.33%
SQ Functional Correctness - RIP	15	57.14%	0.156	0.03	2.7%
SQ Performance Efficiency	7	64.29%	0.5	0.103	2.7%
SQ Resource Utilization	3	64.29%	0.5	0.103	2.7%
Business	18	85.71%	0.33	0.155	2%
SQ Capacity	1	0%	0	0	0%
SQ Compatibility	5	57.14%	0.085	0.01	0%
SQ Maintainability	3	28.57%	0.08	0.007	0%
SQ Portability	5	78.57%	0.5	0.054	0%
SQ Reliability	8	42.86%	0.154	0.016	0%
SQ Usability	6	64.29%	0.16	0.035	0%

Maximum to average impact of criteria sets



Interpret

- In the homogeneous set of 6 cases, 219 criteria:
  - Some widely used criteria have negligible impact and/or variation
  - Some “expensive” criteria are not selective, i.e. not necessarily critical
  - 31 of the 105 mapped criteria are dominated
  - None of the low-impact *sets* is fully dominated
- Some of the criteria that are expensive to measure are in fact insignificant and/or have zero variation
- Summary factors and rankings provide valuable insight

Conclusions

- Formalized semantics increase criterion transparency
  - increases quality and trust
- Set of metrics is helpful for assessing decision criteria in different ways
  - Decision maker: Understand impact of criterion on decisions
  - Researcher: Understand key decision factors in certain scenarios, organisation contexts, etc.
  - Developer: Understand critical success factors of candidates
- Drives improvement of DS environment

DSS improvements

- Filter and shortlist candidates
- Prioritize development of automation
- Rank criteria to measure and cut off data collection
- Approximate non-critical measures
- Remove non-critical criteria
- Recommend criteria based on selectivity and co-occurrence
- Automated suggestion of objectives and criteria based on organisational context
- …

DPUE

- Today is the deadline for choosing topics
- Next week is the **hard** deadline for submitting a concept
  - This concept is part of the grading scheme
  - Send it upfront to your contact point if you want feedback!
- Make sure to get in touch with your contact point
  - Discussion time will be helpful in all cases
- Any questions? Get in touch!

Thank you for your attention!

Questions?

DPUE task 2 questions?

www.ifs.tuwien.ac.at/~becker
www.ifs.tuwien.ac.at/dp/plato