Process Management Plans

Tomasz Miksa, Stephan Strodl, Andreas Rauber
SBA Research, Austria

Agenda

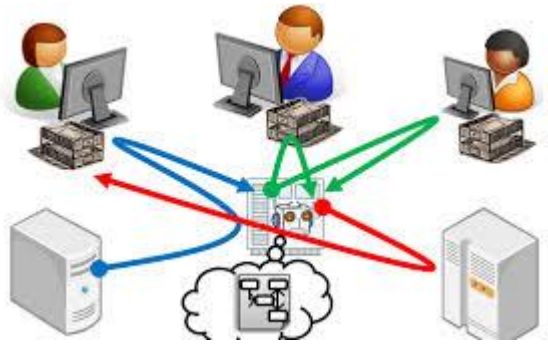
- Process Management Plans (PMPs)
  - **Why** do we need them?
  - **What** are they?
  - **How** to apply them?
- Conclusions and Outlook

Why do we need Process Management Plans?

WHY

eScience and Research Infrastructures

- Scientists exchange
  - facilities
  - resources
  - services
  - datasets
- Research requires
  - special tooling and software
  - processes to
    - transform
    - visualize
    - interpret the data



Research Data

- Data
  - is fuel for research
  - is the result of processes such as
    - capturing
    - pre-processing
    - transformation
    - integration
    - analysis
- Research data quality depends on reproducibility

Reproducibility

- Reproducibility requires
  - well documented research processes
  - precise information on the experiment's environment
- Current studies show very low reproducibility in
  - medicine
  - economy
  - physics
- No incentives for repeating the experiment

Existing approaches

- Data Management Plans
  - are data-centric
  - focus on input and output data
  - collect metadata
  - provide general overview of the experiment
  - are manually created (supported by online tools)
  - applicable to many domains
  - in use

Existing approaches

- Executable papers
  - require controlled environment
  - limited applicability
  - not popular
- Ad hoc actions
  - non standardised
  - limited to an institution or a single domain

Requirements

- We need a domain independent solution to
  - document the experiment data
  - describe precisely the experiment (process)
  - specify the tools, infrastructure and their settings used for the experiment
  - ensure possibility to later verify the experiment

Requirements

- We may not influence the researchers
  - by changing the way they do their research
  - by imposing extra duties on them

What are the Process Management Plans?

WHAT

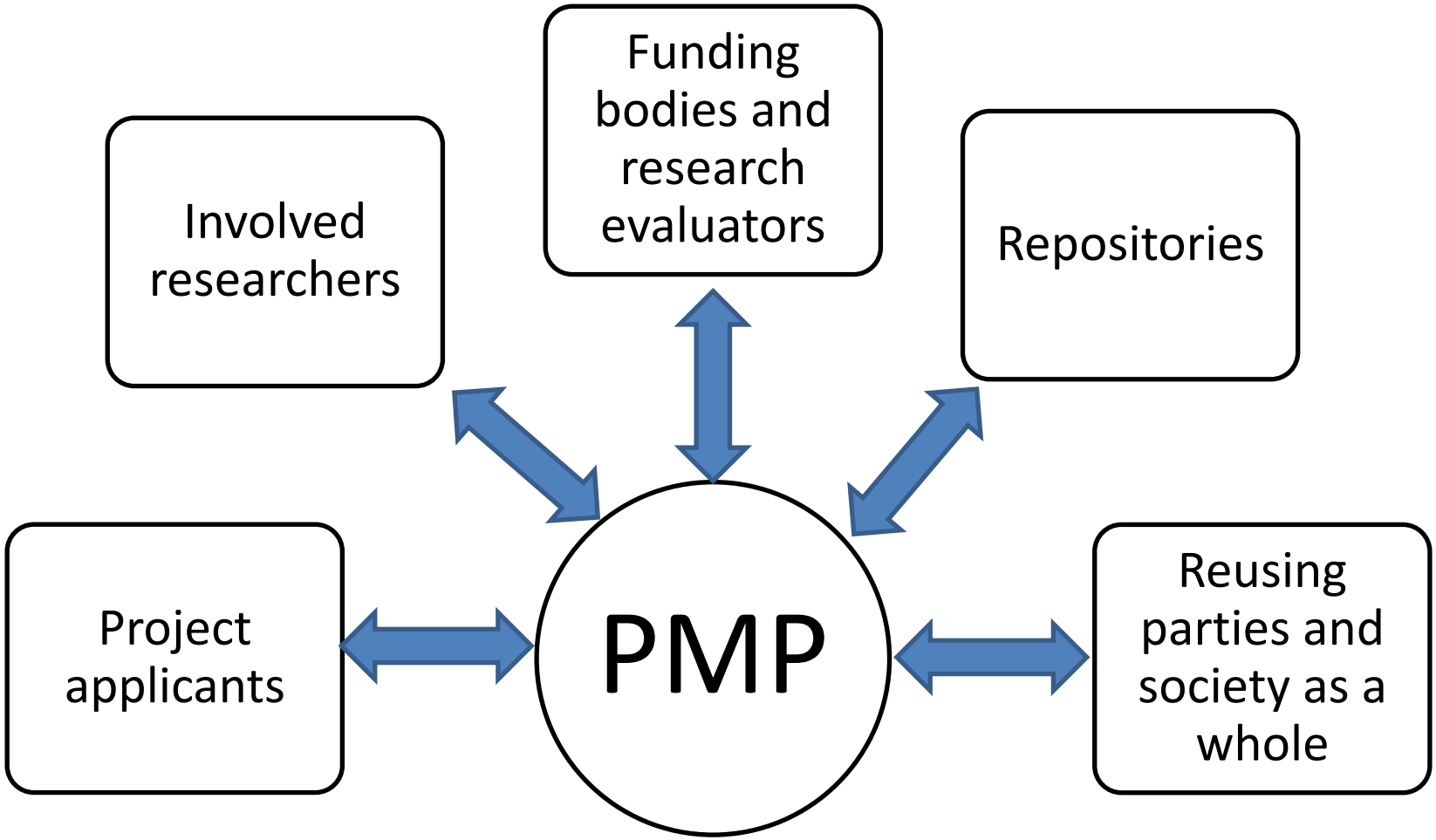
Process Management Plans

- PMPs
  - are process-centric
  - are machine process able
  - have parts generated automatically
  - follow precisely defined schema
  - do not alter the research process
  - complement Data Management Plans

PMPs contents

1. Overview and context
2. Description of process and its implementation
   a. Process description
   b. Process implementation
   c. Data used and produced by process
3. Preservation
   a. Preservation history
   b. Long term storage and funding
4. Sharing and reuse
   a. Sharing
   b. Reuse
   c. Verification
   d. Legal aspects
5. Monitoring and external dependencies
6. Adherence and Review

PMPs stakeholders and benefits

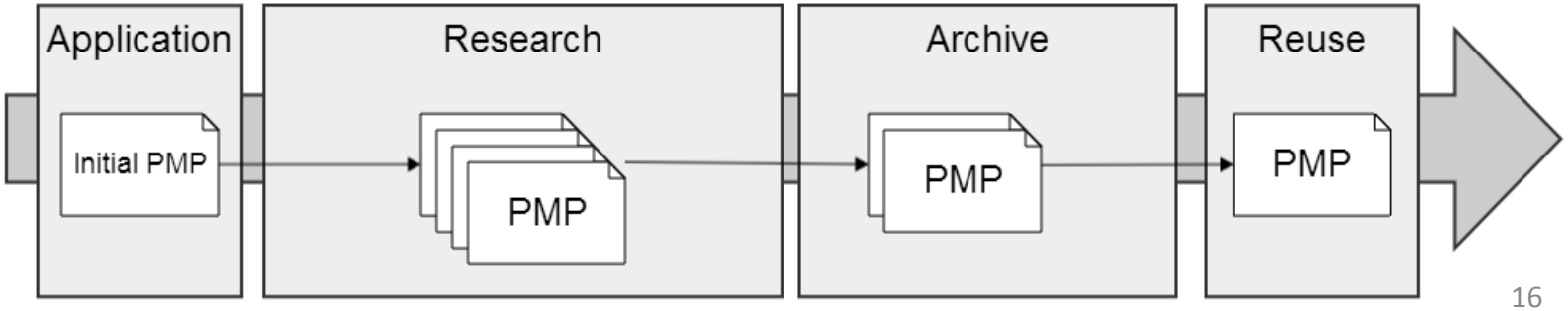


How to apply the Process Management Plans?

HOW

PMPs lifecycle

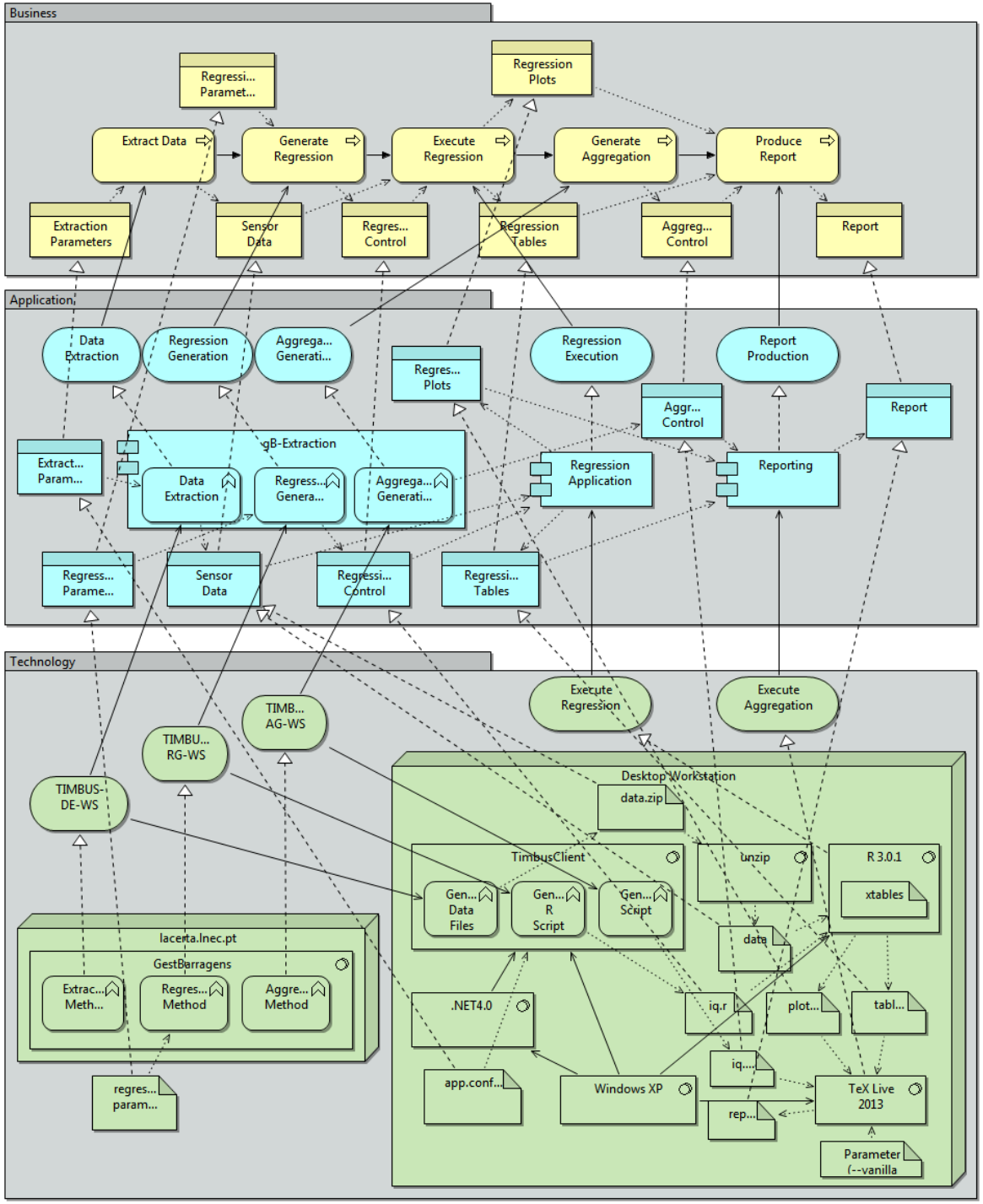
- Updated during lifetime by multiple stakeholders
  - Research
    - created when publishing results, finishing project stage, etc.
  - Archive
    - preservation and monitoring actions
  - Reuse
    - when adjusted for redeployment



PMPs implementation

- Proposed implementation
- Process specific parts of PMP
  - using TIMBUS Context Model
    - is an ontology
    - based on enterprise architecture framework
    - formalizes process description
    - describes process on different levels
    - extendable by domain specific ontologies
    - parts can be generated

TIMBUS
TIMELESS BUSINESS



Business

Application

Technology

PMPs implementation

- Other parts of PMP
  - using existing tools and techniques
- Examples
  - Tools for data management
    - interface for manual provision of data to the PMP
    - collection of information covered by the DMP
      - http://www.dcc.ac.uk/dmponline
  - Research objects
    - to group together artefacts describing the experiment
      - http://www.researchobject.org/

PMPs implementation

- Preservation planning
  - to update the PMPs in the archive if preservation applies
    - http://www.ifs.tuwien.ac.at/dp/plato/intro.html
- Monitoring using SCOUT
  - to monitor external dependencies used by the process and modify the PMP if needed
    - http://www.scape-project.eu/

Process Management Plans summarized

CONCLUSIONS

Conclusions

- Process Management Plans
  - are process-centric view on eScience research
  - complement the data-centric view of Data Management Plans
  - increase the reproducibility of research what enhances the data quality
  - are domain independent
  - do not affect the way the research is performed
  - their implementation is based on existing solutions from data management and preservation community

Outlook

- Development within TIMBUS project
  - Tools for Context Model generation
  - Template specification for manual data acquisition
  - Methods for verification of rerun processes
- Evaluation on use cases from different scientific communities

Thank you!