

Data Management Plans

„My data? Oops....“

GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung

Informationen für Geistes- und Sozialwissenschaftler/innen Förderkriterien für wissenschaftliche Editionen in der Literaturwissenschaft

ANHANG: Basisformat zur digitalen Sicherung von Textdaten aus wissenschaftlichen Editionsprojekten⁴

- Der Text wird in Unicode abgespeichert und mit einem XML-basierten Markup codiert. Dabei sollte dem Modell der Text Encoding Initiative (TEI) in der jeweils aktuellen Version gefolgt werden (z. B. TEI Simple oder DTABf)⁵.
- Eine begründete Entscheidung über die Repräsentation der wesentlichen Strukturelemente und Gliederungsebenen des Textes im Markup werden erwartet. Als Orientierung eignet sich die Sammlung Baseline Encodings in TextGrid⁶ oder Vergleichbares.
- Um die Austauschbarkeit und Nachnutzung zu ermöglichen, werden die projektspezifisch verwendeten XML-Elemente und Attribut-Wert-Paare im TEI-Header dokumentiert.
- Soll auch die Präsentation des Volltextes langfristig gesichert werden, sollte eine geeignete Formatierungssprache gewählt werden (z. B. XSLT, CSS), die die Unabhängigkeit von spezieller Software weitgehend sicherstellt.
- Die abgelegten Daten sollten im Idealfall der Version des Textes entsprechen, die auch der finalen⁷ Fassung der Veröffentlichung entspricht. Abweichungen sind zu begründen, etwa bei der Auswahl einer bestimmten Textfassung.
- Die Texte werden zur Archivierung in einem institutionellen oder fachlichen Repository abgelegt, das die langfristige Verfügbarkeit und Zitierfähigkeit
- durch persistente Identifikatoren
- gewährleisten kann. Das Repository sollte durch ein DINI-Zertifikat⁸ oder ein vergleichbares Qualitätssiegel ausgezeichnet sein.
- Die elektronische Ablage erfolgt ausdrücklich unter einer Lizenz, die eine freie Nachnutzung zu wissenschaftlichen Zwecken erlaubt. ⁹ Dem Antrag sollte eine Bestätigung der Institution beiliegen, welche die Langzeitarchivierung der Textdaten gewährt.

Informationen für Geistes- und Sozialwissenschaftler/innen Förderkriterien für wissenschaftliche Editionen in der Literaturwissenschaft

ANHANG: Basisformat zur digitalen Sicherung von Textdaten aus wissenschaftlichen Editionsprojekten⁴

- Der Text wird in Unicode abgespeichert und mit einem XML-basierten Markup codiert. Dabei sollte dem Modell der Text Encoding Initiative (TEI) in der jeweils **aktuellen Version** gefolgt werden (z. B. TEI Simple oder DTABf) .
- Eine begründete Entscheidung über die Repräsentation der wesentlichen Strukturelemente und Gliederungsebenen des Textes im Markup werden erwartet. Als Orientierung eignet sich die Sammlung Baseline Encodings in TextGrid oder Vergleichbares.
- Um die **Austauschbarkeit und** Nachnutzung zu ermöglichen, werden die projektspezifisch verwendeten XML-Elemente und Attribut-Wert-Paare im TEI-Header dokumentiert.
- Soll auch die Präsentation des Volltextes **langfristig gesichert** werden, sollte eine geeignete Formatierungssprache gewählt werden (z. B. XSLT, CSS), die die **Unabhängigkeit von spezieller** Software weitgehend sicherstellt.
- Die abgelegten Daten sollten im Idealfall der Version des Textes entsprechen, die auch der finalen **Fassung der Veröffentlichung** entspricht. Abweichungen sind zu begründen, etwa bei der Auswahl einer bestimmten Textfassung.
- Die Texte werden zur Archivierung in einem institutionellen oder fachlichen Repository abgelegt, das die langfristige Verfügbarkeit und Zitierfähigkeit
- durch **persistente Identifikatoren**
- gewährleisten kann. Das Repository sollte durch ein DINI-Zertifikat oder ein vergleichbares **Qualitätssiegel** ausgezeichnet sein.
- Die elektronische Ablage erfolgt ausdrücklich unter einer Lizenz, die eine **freie Nachnutzung** zu **wissenschaftlichen** Zwecken erlaubt. Dem Antrag sollte eine Bestätigung der Institution beiliegen, welche die Langzeitarchivierung der Textdaten gewährt.

Informationen für Geistes- und Sozialwissenschaftler/innen Förderkriterien

- **Use XML format such as DTABf:** The data format for historic data developed by the CLARIN centre at the BBAW in Berlin
- **Independent of specific software:** CLARIN uses open interfaces
- **Version of the publication:** CLARIN supports
- **Deposit in institutional or subject specific repository:** The CLARIN centres have repositories
- **Using persistent identifiers:** CLARIN supports ISO 24619:2011
- **Certified repositories:** CLARIN repositories: DSA + centre assessment
- **Free use for research and teaching:** CLARIN-Supports authentication for the academic community

ten⁴

p codiert. Dabei sollte dem
gt werden (z. B. TEI Simple

strukturelemente und
eignet sich die Sammlung

ektspezifisch verwendeten

eine geeignete
gkeit von spezieller

, die auch der finalen
etwa bei der Auswahl einer

epositorium abgelegt, das

ein vergleichbares

e Nachnutzung zu
ler Institution beiliegen,

- Empfehlung 7: Sicherung und Aufbewahrung von Primärdaten

Primärdaten als Grundlagen für Veröffentlichungen sollen auf haltbaren

10 years und gesicherten Trägern in der Institution, wo sie

entstanden sind, zehn Jahre lang aufbewahrt werden.

- Empfehlung 7: Sicherung und Aufbewahrung von Primärdaten
- Primärdaten als Grundlagen für Veröffentlichungen sollen gesicherten, dauerhaft verfügbaren und

**Long lasting, enduring
and safe media**



- Processes
 - Depositing
 - Archiving
 - Creation of metadata
 - Evaluation
- Technology
 - Operation
 - Backup
 - Persistency
 - Standards, e.g. OAIS
- Rights and legal
 - Depositors' agreement (contracts !)
 - Access control
 - Access on hold
- ...

- http://www.dfg.de/foerderung/antragstellung_begutachtung_entscheidung/antragstellerende/antragstellung/nachnutzung_forschungsdaten/

Reuse

Type of research data

Quality assessment

Long time archiving

Data type

Discipline specific standards

***Appropriate
repositories***

***Rights of third
parties***

Timeframe

- Bereits in die Planung eines Projekts sollten Überlegungen einfließen, ob und welche der aus einem Vorhaben resultierenden Forschungsdaten für andere Forschungskontexte relevant sein können und in welcher Weise diese Forschungsdaten anderen Wissenschaftlerinnen

Data management plan

- und Wissenschaftlern zur Nachnutzung zur Verfügung gestellt werden können. In einem Antrag sollten die Antragstellenden daher ausführen, welche Forschungsdaten im Verlauf eines wissenschaftlichen Forschungsvorhabens entstehen, erzeugt oder ausgewertet werden.
http://www.dfg.de/foerderung/antragstellung_begutachtung_entscheidung/antragstellende/antragstellung/nachnutzung_forschungsdaten/

- “Already during the planning phase of a project it should be considered if and which research data resulting from the project could be relevant for other research contexts and in which way these bits of data could be made accessible to other researchers for reuse. Within an application, applicants should described in detail which research data will be created in the research project, how they are generated or analysed.”
- http://www.dfg.de/foerderung/antragstellung_begutachtung_entscheidung/antragstellende/antragstellung/nachnutzung_forschungsdaten/

- Data management plan
 - created
 - submit
 - implement
- Formal requirements of German funders:
 - DFG
 - BMBF
 - VW foundation
 - :

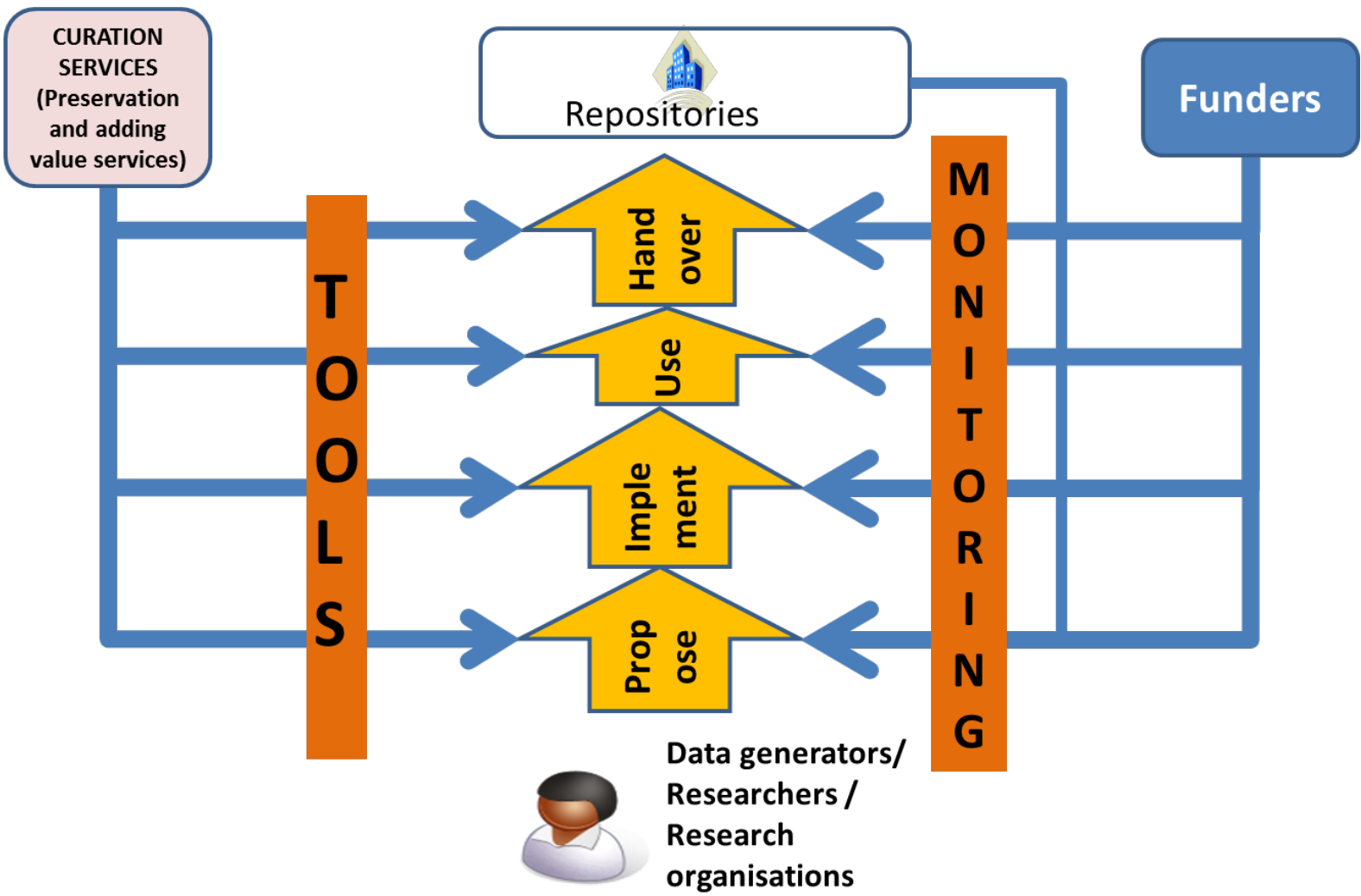


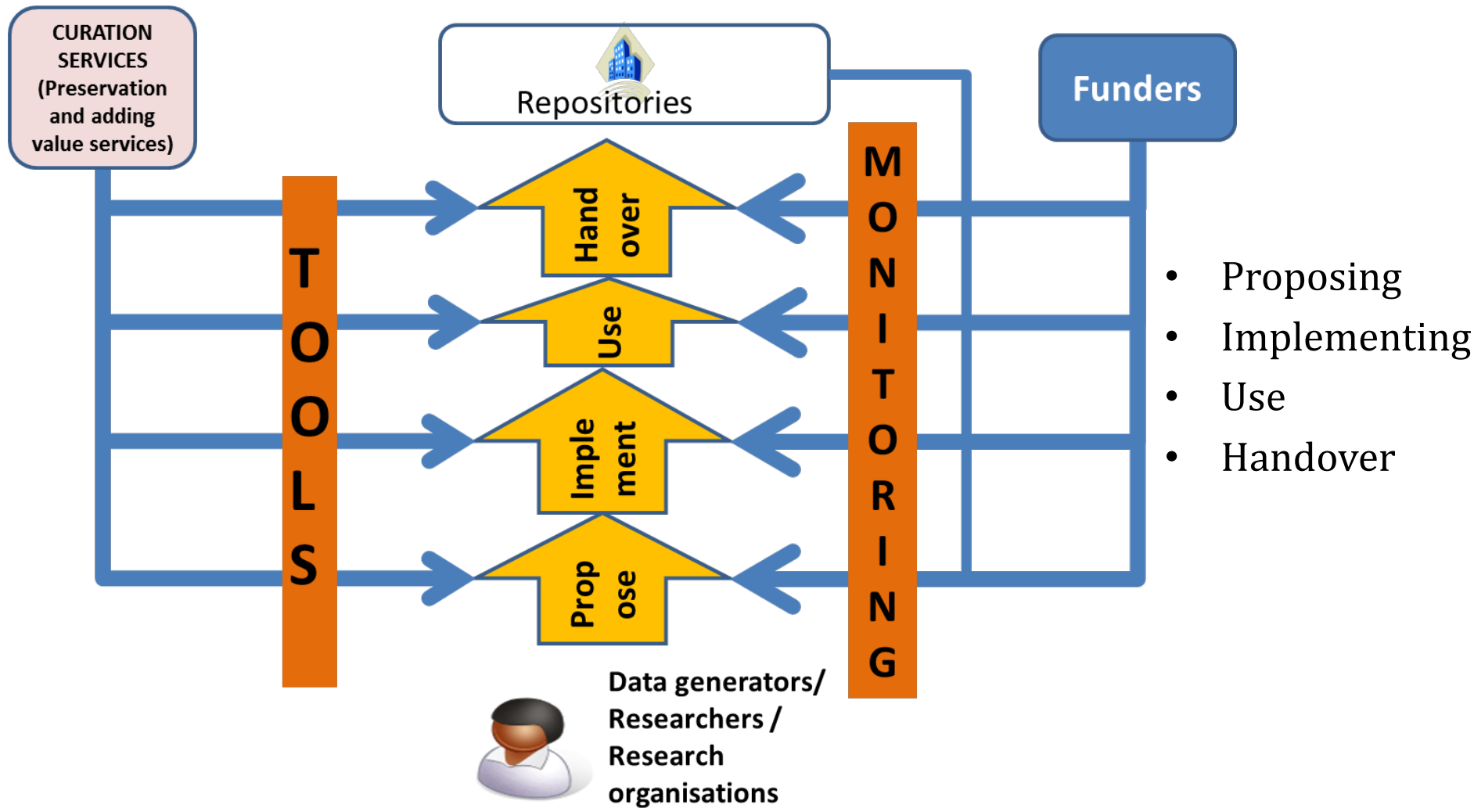
- #\$\$%&*! (Disgusting)
 - Each Research Council/Stakeholder
 - Digital Curation Centre (DCC). Funders' data plan requirements. See <http://www.dcc.ac.uk/resources/data-management-plans/funders-requirements> .
 - http://www.dcc.ac.uk/sites/default/files/documents/resource/DMP/DMP_Checklist_2013.pdf
 - http://www.dcc.ac.uk/sites/default/files/documents/data-forum/documents/docs/DCC_Checklist_DMP_v3.pdf
 - Research Data Management Policy at the University of Edinburgh. See <http://www.ed.ac.uk/files/imports/fileManager/ResearchDataManagementPolicy20150313.pdf>
 - Funder Requirements at the Digital Curation Centre (DCC). See <http://www.dcc.ac.uk/resources/data-management-plans/funders-requirements>
 - Checklist for a Data Management Plan. v.4.0, Edinburgh: Digital Curation Centre, 2013. Available online: <http://www.dcc.ac.uk/resources/data-management-plans> .
- Switzerland
 - Guidelines at the Swiss National Science Foundation. See http://www.snf.ch/sitecollectiondocuments/allg_reglement_e.pdf, Article 44(b).
 - Guidelines for Research Integrity at the ETH Zurich. See <https://www.ethz.ch/content/dam/ethz/main/research/pdf/forschungsethik/Broschure.pdf> .
- EU
 - Guidelines on Data Management in Horizon 2020 (Version 2.1). European Commission, Directorate-General for Research & Innovation. See http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf , July 2016.

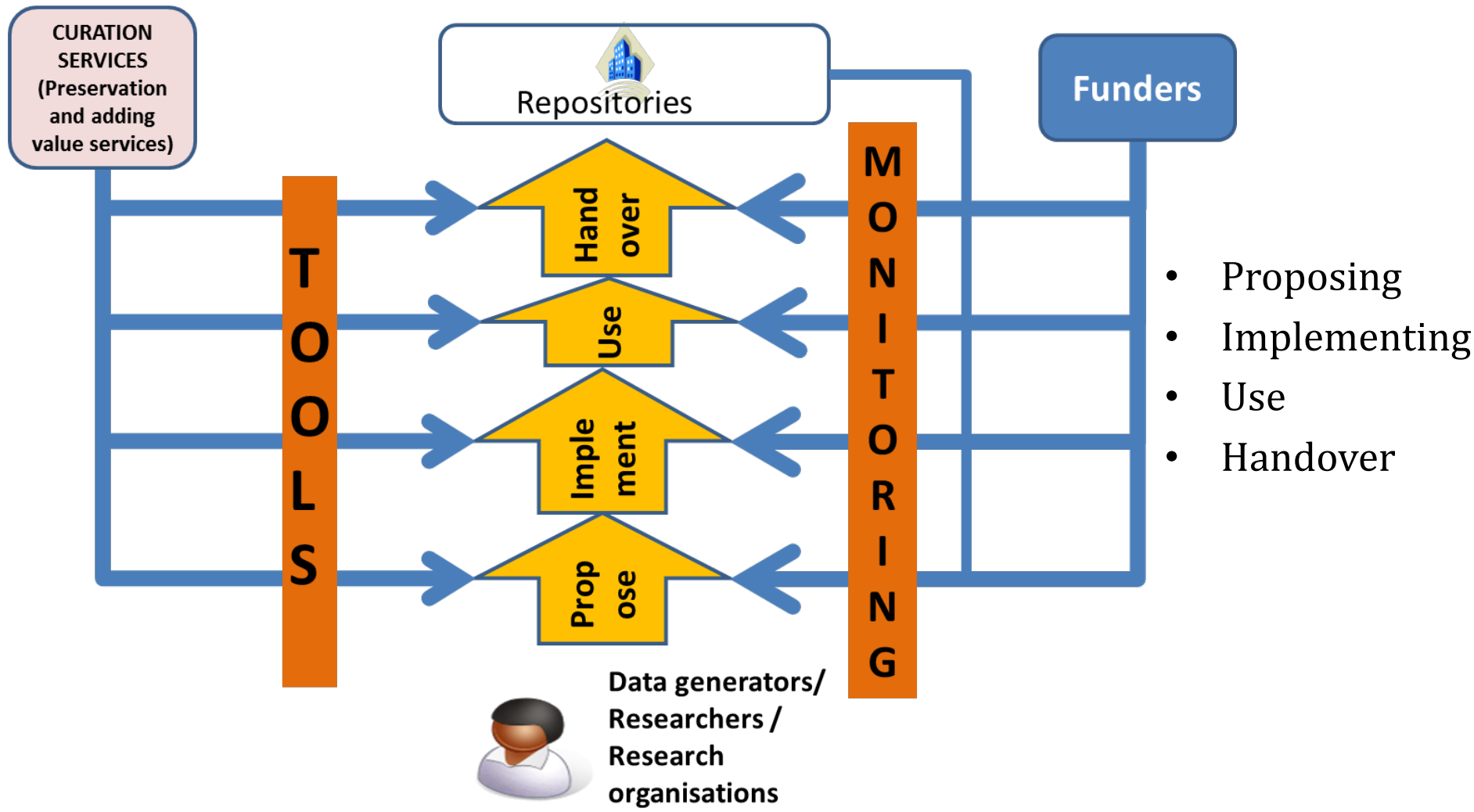
- http://www.dcc.ac.uk/sites/default/files/documents/resource/DMP/DMP_Chcklist_2013.pdf
- For data management, one checklist has the following groups of questions:
 - Administrative data
 - Data collection
 - Documentation and metadata
 - Ethics and legal compliance
 - Storage and backup
 - Selection and preservation
 - Data sharing
 - Responsibilities and resources
 - What do you think is meant with these classes of information?
- Task: read the checklist
 - Which of the questions comes as a surprise?
 - Which question do you think is the hardest to answer? For whom?

- It depends
 - In general: non proprietary
 - Open specifications
 - Non binary
 - Non compressed
 - Compromises necessary
- Some hints: <http://clarin-d.net/en/language-resources-and-services/user-guide> chapter 6

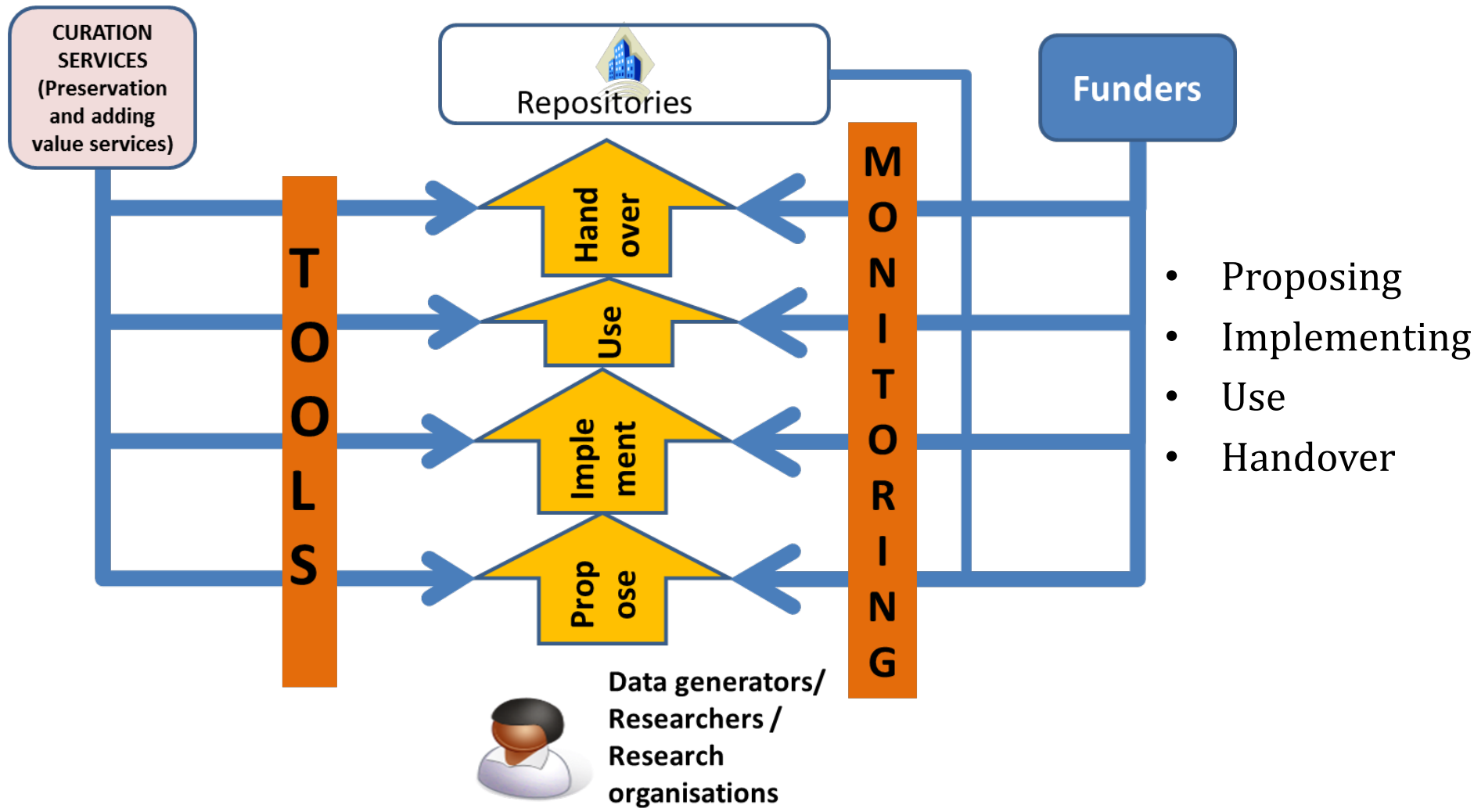
- Terminology of the Research Data Alliance (RDA)
- Working group: ADMP
- See <https://rd-alliance.org/group/active-data-management-plans/wiki/admp-scenarios.html>



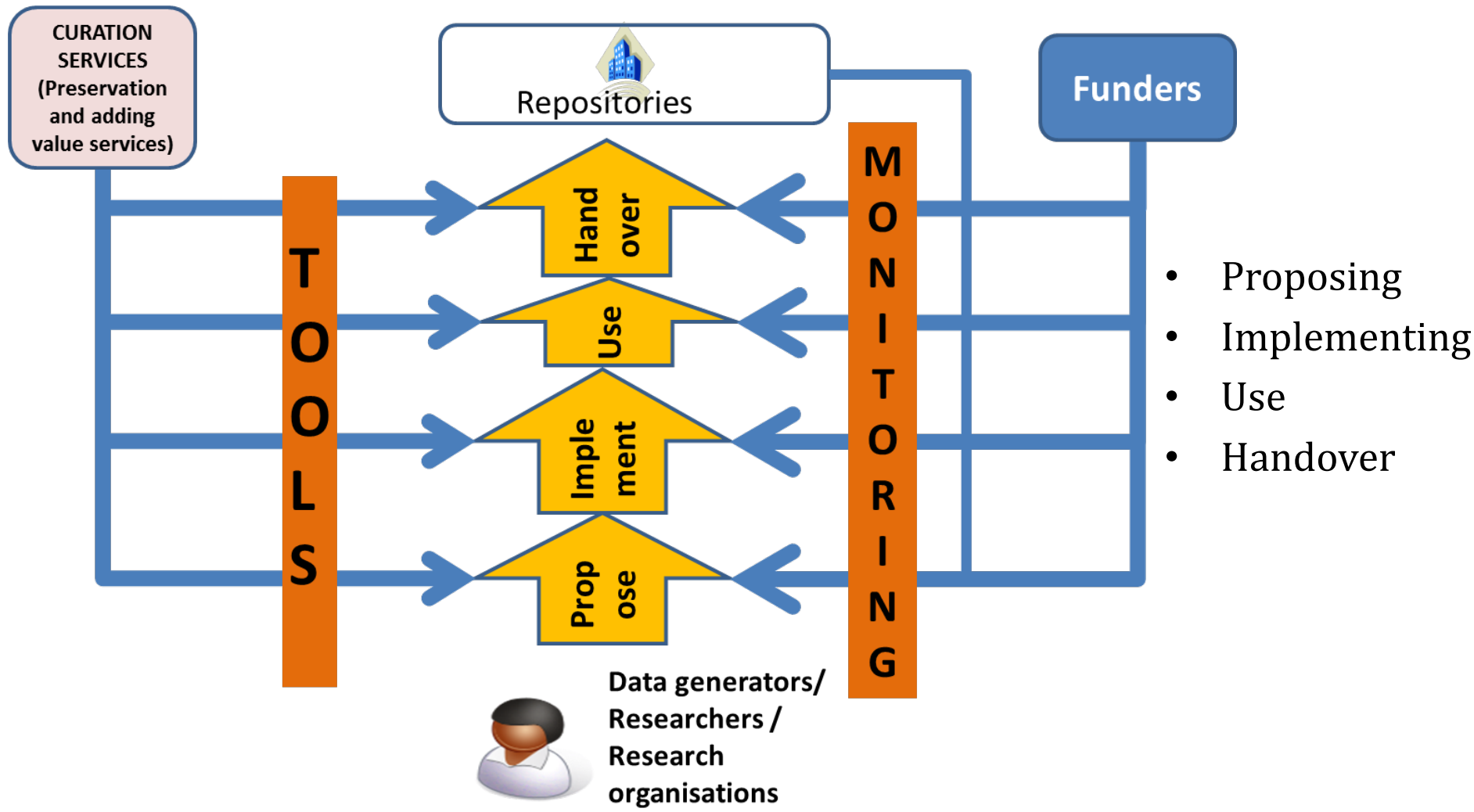




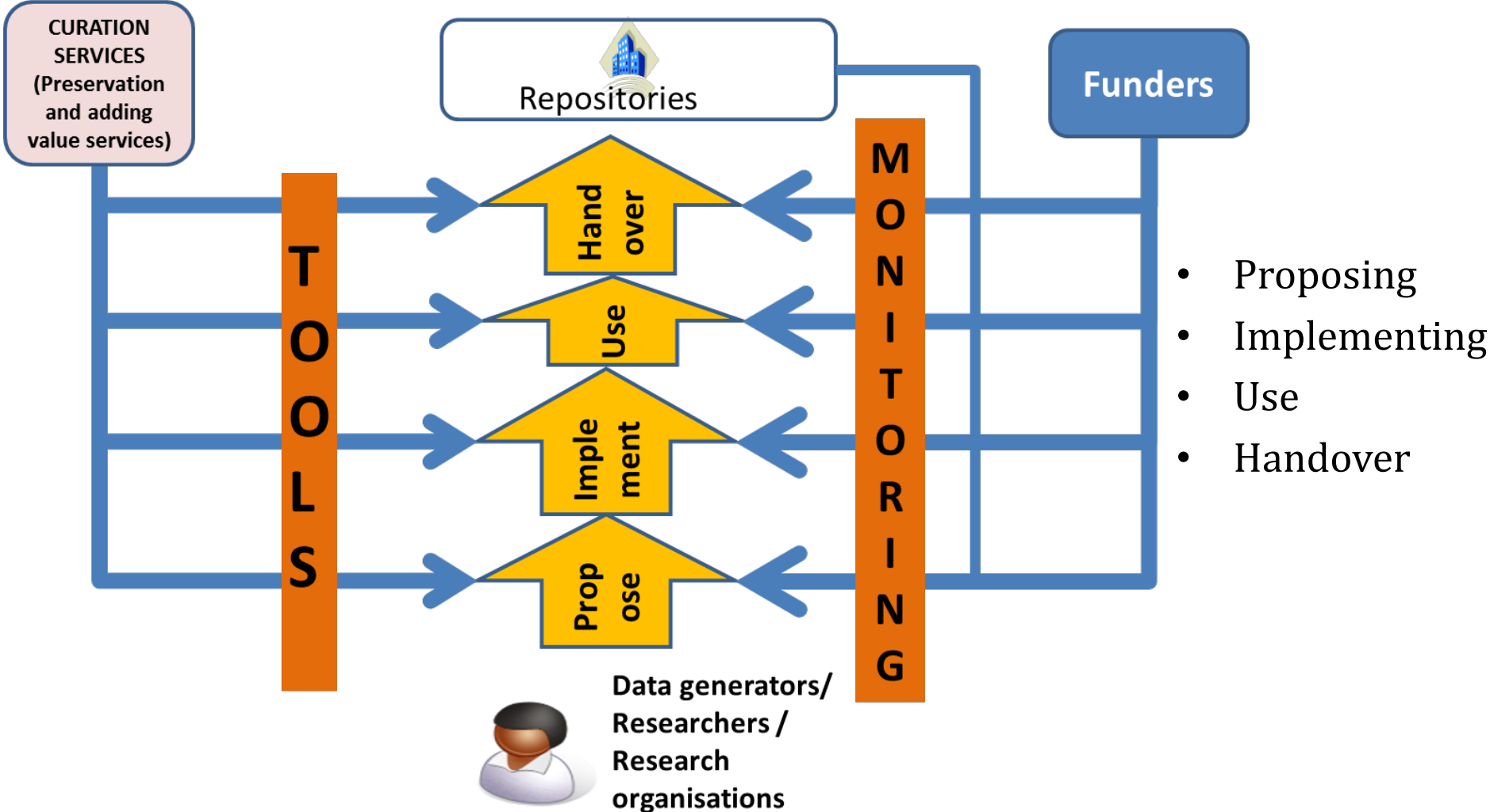
- Project application
 - Research question
 - Sources
 - CV
 - Publications
 - Time line
 - State of the art
 - Previous work
 - Collaboration with others
 - Data Management Plan
 - Defining task
 - Responsibilities
 - Deadlines



- Which options do you think there are to follow up the implementation of the DMP?
- Who is responsible?
- Where are the problems?
- What is your experience WRT this kind of bureaucracy?



- Use of tools
- Publishing about data
- Citing
- Restricted access to tools



- When is an appropriate time for handing over data?
- What can be made public?
- Which criteria?

- Digital Curation Centre (DCC). Funders' data plan requirements. See <http://www.dcc.ac.uk/resources/data-management-plans/funders-requirements>.
- http://www.dcc.ac.uk/sites/default/files/documents/resource/DMP/DMP_Checklist_2013.pdf
- http://www.dcc.ac.uk/sites/default/files/documents/data-forum/documents/docs/DCC_Checklist_DMP_v3.pdf
- Research Data Management Policy at the University of Edinburgh. See <http://www.ed.ac.uk/schools-departments/information-services/research-support/data-management/data-management-planning>.
- Funder Requirements at the Digital Curation Centre (DCC). See <http://www.dcc.ac.uk/resources/data-management-plans/funders-requirements>.
- Checklist for a Data Management Plan. v.4.0, Edinburgh: Digital Curation Centre, 2013. Available online: <http://www.dcc.ac.uk/resources/data-management-plans>.
- Guidelines at the Swiss National Science Foundation. See http://www.snf.ch/sitecollectiondocuments/allg_reglement_e.pdf, Article 44(b).
- Guidelines for Research Integrity at the ETH Zurich. See <https://www.ethz.ch/content/dam/ethz/main/research/pdf/forschungsethik/Broschure.pdf>.
- Guidelines on Data Management in Horizon 2020 (Version 2.1). European Commission, Directorate-General for Research & Innovation. See http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf, February 2016.
- <https://rd-alliance.org/group/active-data-management-plans/wiki/admp-scenarios.html>
- <http://clarin-d.net/en/language-resources-and-services/user-guide>

- <https://clarin-d.net/en/preparation/data-management-plan>
- Alternative: www.clarin-d.net
 - „Preparation “
 - → Data management plan