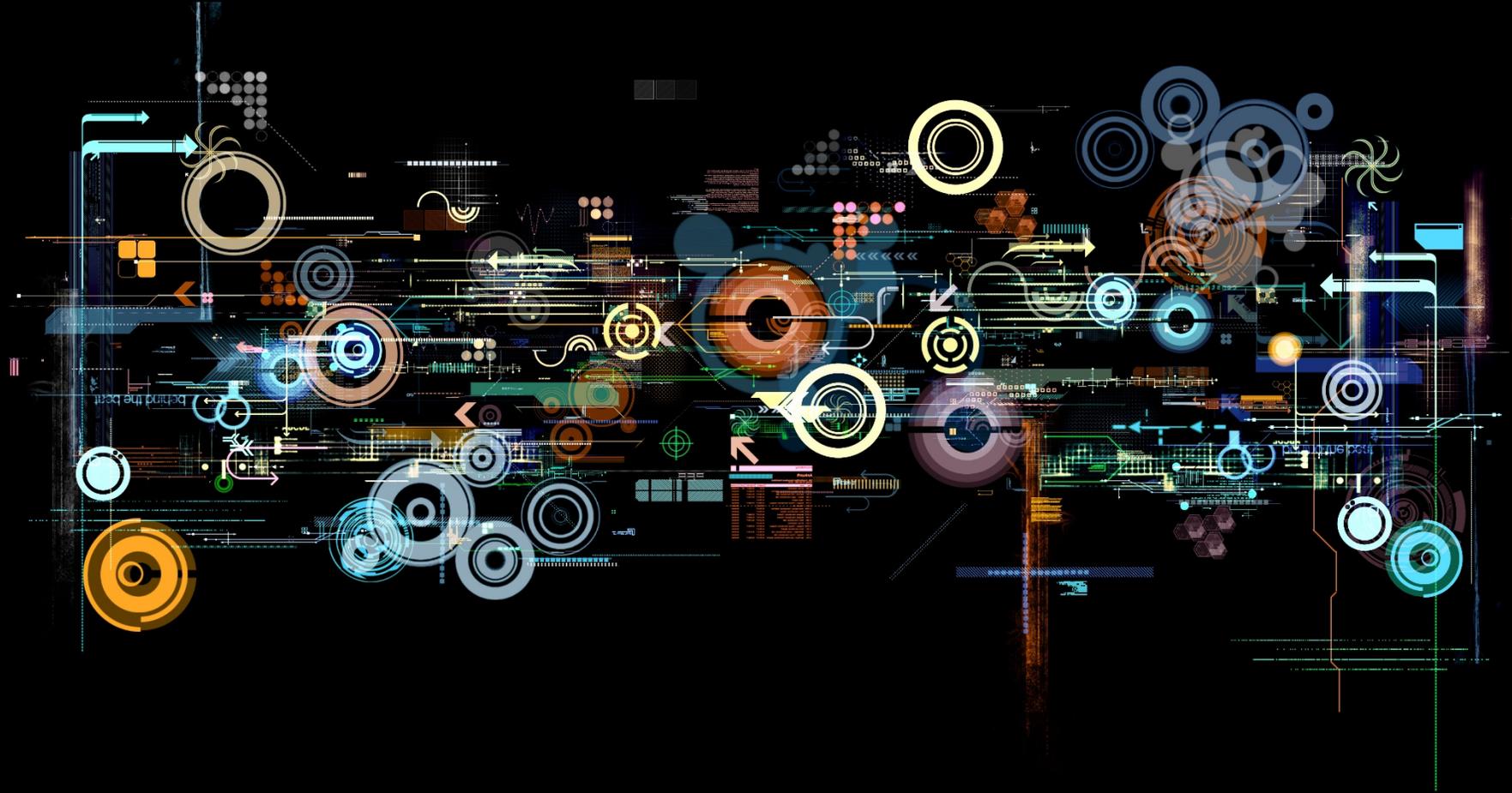


NEW DATA SPACES FOR THE SOCIAL SCIENCES

AN INFRASTRUCTURE PRIORITY PROGRAMME



NEUEN DATENRÄUME FÜR DIE SOZIALWISSENSCHAFTEN

DFG Infrastrukturschwerpunktprogramm

- **DFG:**
 - Eckard Kämper
- **Programme Committee**
 - Cordula Artelt, LfBi Bamberg (Speaker)
 - Stefan Liebig, FU Berlin
 - Corinna Kleinert, LfBi Bamberg
 - Alexander Mehler, GU Frankfurt
 - Reinhard Pollak, GESIS Mannheim



KSWD, 27.03.2023 | 17:45 – 18:15 Uhr

www.new-data-spaces.de

Neue Datenräume für die Sozialwissenschaften

Merkmale koordinierter Förderung in Schwerpunktprogrammen

- ▶ Grundidee der Förderung: ortsübergreifende, interdisziplinäre Kooperation
- ▶ Struktur: ca. 10-15 ortsverteilte Einzelprojekte, Zentralprojekt
- ▶ Grundlage: ein aus der Wissenschaft entwickeltes und vorgeschlagenes Konzept
- ▶ Die Konzepte für Schwerpunktprogramme werden begutachtet, der Senat entscheidet über die Einrichtung
- ▶ Schwerpunktprogramme werden für 2 x 3 Jahre eingerichtet
- ▶ Auf dieser Grundlage ruft die DFG zur Einreichung von Einzelanträgen auf
- ▶ Eine Begutachtungsgruppe wählt aus den Anträgen aus
- ▶ Sonderform Infrastrukturschwerpunktprogramm: Verlängerung über die ersten 6 Jahre hinaus möglich
- ▶ Das bedeutet: eine außergewöhnliche Gelegenheit, die Transformation der sozialwissenschaftlichen Dateninfrastruktur zu gestalten

Neue Datenräume für die Sozialwissenschaften (SPP 2431)

Zeitplan

- ▶ 7. Dezember 2022: Entscheidung des Senats über die Einrichtung des Infrastrukturschwerpunkts
- ▶ 14. März: Ausschreibung (Information für die Wissenschaft Nr. 20)
https://www.dfg.de/foerderung/info_wissenschaft/2023/info_wissenschaft_23_20/index.html
- ▶ 27. März: 1. Vorstellung des Konzepts auf der Konferenz für Sozial- und Wirtschaftsdaten
- ▶ 12. Mai: vorbereitender Workshop in Bonn (Anmeldung per Projektidee bis zum 17. April; Anmeldung bei Koordinatoren; die DFG lädt ein; Teilnahme keine Voraussetzung für die Antragstellung)
- ▶ 20. September: Antragsfrist
- ▶ Winter 2023: Begutachtung durch eine internationale Begutachtungsgruppe
- ▶ Förderentscheidungen durch die DFG im Frühjahr 2024
- ▶ Programmstart: Sommer 2024



Vielen Dank für Ihre Aufmerksamkeit

Weitere Informationen

- ▶ zur DFG: <https://www.dfg.de>
- ▶ zum Förderatlas: <https://www.dfg.de/foerderatlas>
- ▶ zu allen geförderten Projekten: <https://www.dfg.de/gepris>
- ▶ zu den deutschen Forschungseinrichtungen: <https://gerit.org>
- ▶ zu historischen DFG-Projekten: <https://gepris-historisch.dfg.de>

PROGRAMME COMMITTEE

- Prof. Dr. C. Artelt** Leibniz Institute for Educational Trajectories (LifBi), Bamberg, Director & University Professor, Educational Research
- Prof. Dr. C. Kleinert** Leibniz Institute for Educational Trajectories (LifBi), Bamberg, Associate Director & University Professor, Sociology
- Prof. Dr. S. Liebig** Freie Universität Berlin, University Professor, Empirical Social Structure Analysis and Survey Methodology
- Prof. Dr. A. Mehler** Goethe Universität, Frankfurt, Head of Text Technology Lab, University Professor, Computational Humanities/Text Technology
- Prof. Dr. R. Pollak** GESIS - Leibniz Institute for the Social Sciences, Mannheim, Scientific Director Department DRS & University Professor, Sociology



WHY DO WE NEED AN INFRASTRUCTURE PRIORITY PROGRAMME



- **Aim:** Make a lasting impact on social sciences research and the survey and data landscape
- **Belief:** “*Not getting behind*” - Data landscapes as a central asset of the future | Counteracting compartmentalization: Complex methodology and tasks cannot be done and solved by a single discipline
- **Mission and Scientific gain:** Understand and push forward issues specific to new data types and their linkage from different perspectives and develop strategies for incorporating these data into different types of robust designs of social sciences
- **Deliverables:** Create and adapt multimodal data acquisition and linking methods in ways that adhere to the design structures, measurement rigor and ethical protections of social science | Sustainable research data and survey landscape

WHY DO WE NEED NEW DATA SPACES?

Survey data: Large-Scale Panel or Trend Studies in the Social Sciences, e.g.

SOEP | NEPS | FReDA | FGZ | PASS | ALLBUS | ESS | PISA | TIMMS | PIAAC

Strengths: e.g. Known sample base | Psychometric properties

Geo-referenced data

...

Big data

...

Register data

...

Synthetic data

...

New types of data acquisition, e.g.

Virtual reality | NLP | Augmented reality | Ecological momentary assessment

Strengths: e.g. Agility | Dynamics | Density

Strengths of integration and synergies

- Validity gains & new constructs
- Richer data & increased efficiency

NEW DATA SPACES AS EMERGING FIELDS OF RESEARCH

- The social sciences need to **unlock opportunities of new data spaces** to expand our knowledge about major social challenges of the coming decades
 - digitization, climate change, growing diversification, pandemics, and war-induced societal changes
- Systematic and **far-sighted social science research needs to rely on**
 - **new forms of data** | e.g. digital trace data and data from administrative processes
 - **new forms of data acquisition** | e.g. virtual interviews and augmented reality
 - **new quality measures for method inherent bias** | e.g. machine learning approaches

AIMS

- Driving a surge in innovation **by improving, enhancing, and combining existing data sources**, and data that have only recently become available in order **to develop new data spaces** for the social sciences

SCOPE

The InfPP contributes to this aim by

- **bundling the skills, knowledge, and expertise** in different **fields of empirical social research and computer science,**
- **providing the means** to test new methods and procedures **of data generation,**
- **developing convergence methodologies and best practice examples** of data sources, data collection, and data integration while increasing the breadth of tools and data available to social science research,
- **enhancing data quality** and **anchoring innovations** in data collection programs

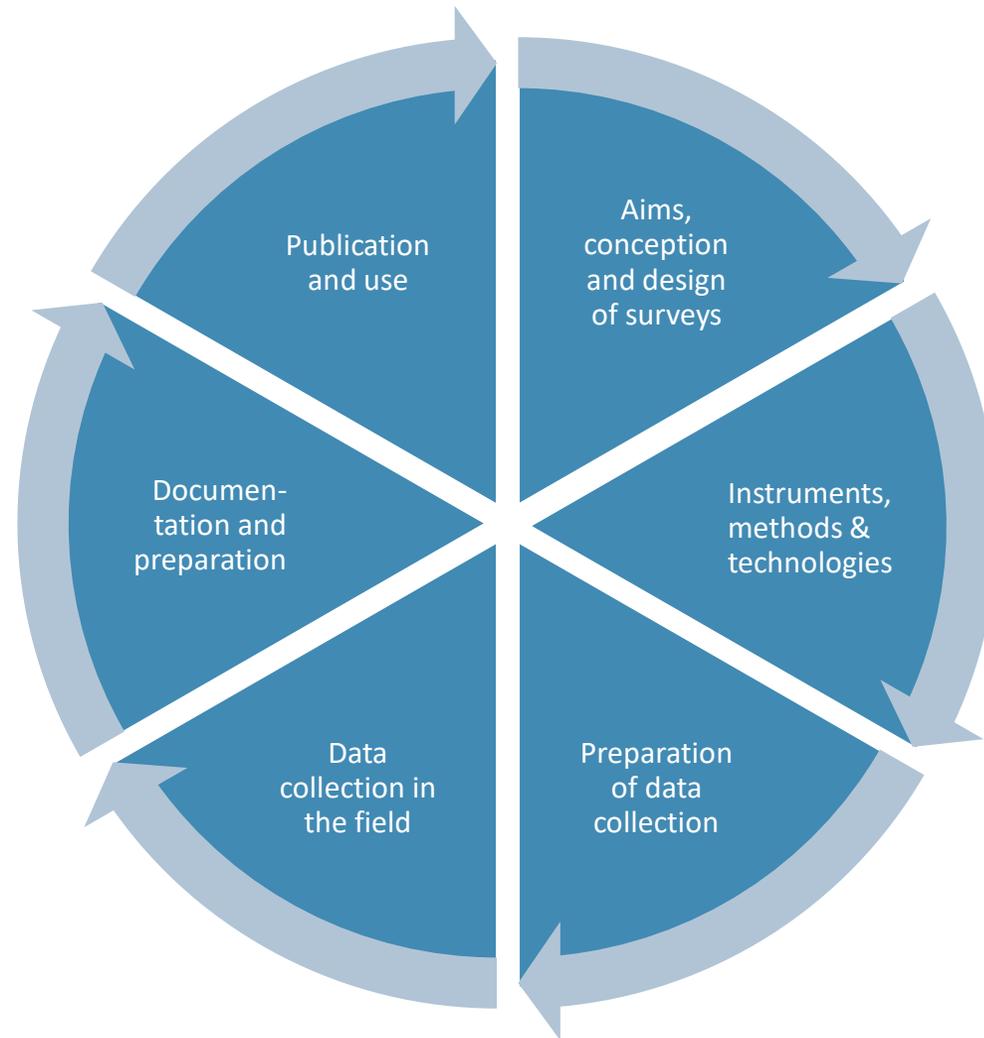
NEW DATA SPACES AS EMERGING FIELDS OF RESEARCH

**Research Area 1:
Data exploration
and integration**

- Survey data
- Register data
- Geo-referenced data
- Big data
- Synthetic data

**Research Area 4:
Multimodal data
acquisition**

- Online surveys and assessment
- App-based surveys
- Virtual interviews and interview bots
- Augmented reality



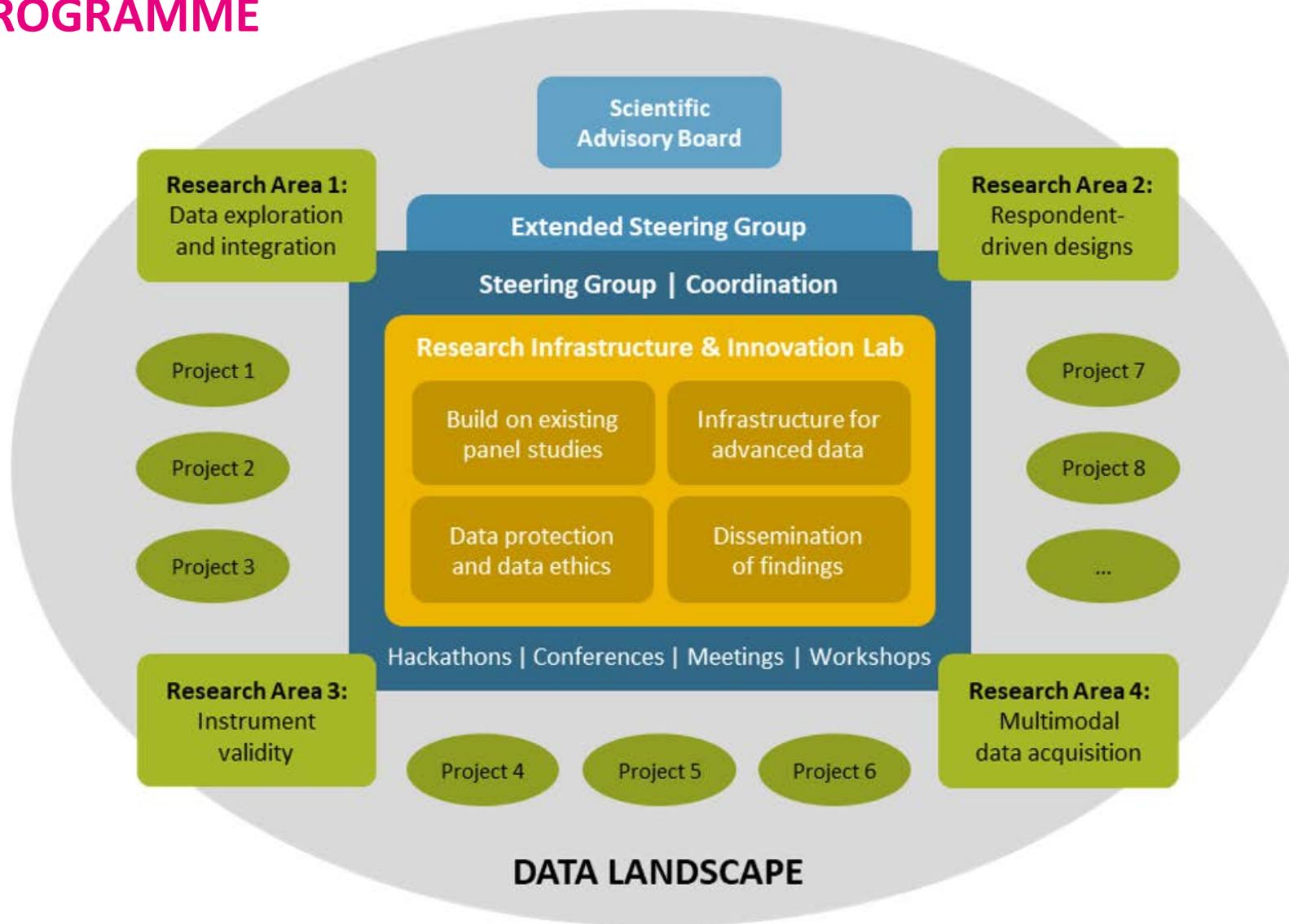
**Research Area 2:
Respondent-driven
designs**

- Recruitability
- Hard-to-reach and hidden populations
- Responsive designs

**Research Area 3:
Instrument
validity**

- Method-inherent bias
- Generalizability
- Measuring new constructs

SUBJECT, ORGANIZATION, AND STRUCTURE OF THE INFRASTRUCTURE PRIORITY PROGRAMME



Unique (successful and supportive) environment for the InfPP's projects

- **Build on existing panel studies** (e.g. large-scale panel studies)
- **Research-driven infrastructure for advanced survey-related data** (e.g. that will provide unique and innovative services for the projects within InfPP)
- **Dissemination and diffusion of results** (e.g. for possible implementations in different types of studies, collaborations, related exchanges within the InfPP)
- **Data protection and ethics** (e.g. with regard to the linking of different data types)

BUILD ON EXISTING PANEL STUDIES

Take advantage of existing data collection infrastructures

➤ **GESIS Panel Infrastructure**

- GESIS Panel Population Sample
- GESIS Panel.plus Digital Behavioral Data Sample



➤ **SOEP Innovation Sample**

- SOEP Innovation Sample Household Survey



➤ **NEPS next**

- NEPS next Add-on-minutes (Outcome CATI)
- NEPS next Add-on-study (linkage with NEPS data)
- NEPS next Add-on-online survey (via LIfBI)



Terms and conditions for the use of data collection infrastructures

- **Data collection infrastructures with specific features**
 - Target population
 - Sampling design
 - Frequency of data collections
 - Data types
 - Survey modes
 - etc.

- **Data collection costs may need to be included in DFG grant proposal**

- **Check website for details and get in touch in program coordinators in advance**

INFRASTRUCTURE FOR ADVANCED DATA | EXPANDING NETWORKS

Core aim of the InfPP

- is to develop and maintain a **research infrastructure for prospective projects**
- by a **multi-stage strategy**, by which the development of new approaches based on new data will be driven, scaled, and evaluated as the InfPP progresses

Maximizing Synergies

- **development and data exploration** activities will be **research-driven** and **oriented to the need of the InfPP projects** for which there are currently no or insufficient offerings and which are beyond the offerings of the existing NFDIs
- from this perspective, the InfPP will strive for **close cooperation with relevant NFDIs**, e.g. Text+, KonsortSWD, BERD, NFDI4DS

INFRASTRUCTURE FOR ADVANCED DATA | ESTABLISHING PROCESSES

To achieve these aims, the InfPP will carry out **six subtasks**

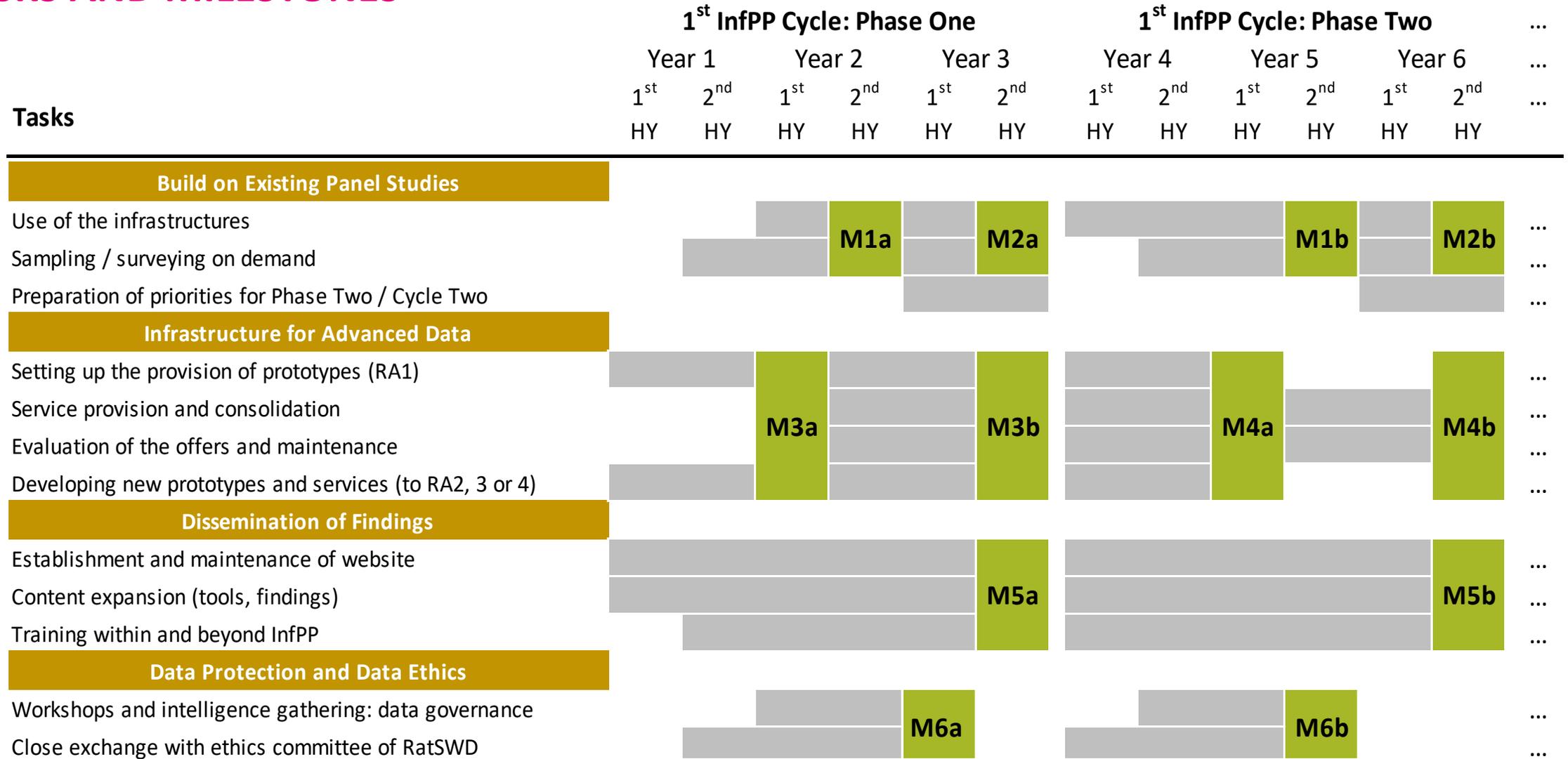
1. **Multimodal survey data acquisition:** collecting behavioral, symbolic, time- & geo-referenced data
2. **Project coherence via data management & data reuse:** document, network and process data for training ML systems
3. **Data linkage, aggregation, and annotation:** generating complex survey data types and their correlations; data annotation and active learning
4. **Automatic survey data analysis:** multi-service platform along with new AI methods for detecting, classifying, aggregating, grouping, linking, and analyzing advanced survey data
5. **Authority management:** multi-user platform that allows researchers to work in different roles or research groups on the same data, protected at the data and tool level
6. **Hub technology:** maintenance and ongoing development of the infrastructure, accompanied e.g. by joint projects, workshops, hackathons

INFRASTRUCTURE FOR ADVANCED DATA | EVOLVING POSSIBILITIES

To meet the needs of the research projects while providing a spectrum of new survey research methods, **spaces of opportunity** are created in a multi-level sequence

- **first stage:** generic prototypes services will be developed to open up new data- and method-related horizons in survey research
- **second stage:** adaption of the infrastructure and the InfPP projects will be able to influence its development based on their arising requirements
- workshops to **continuously consolidate and balances** these objectives
- **assessment of the performance** of the infrastructure towards the end of the third year for **setting research priorities** of the second phase of the InfPP

TASKS AND MILESTONES



HOW WE WORK TOGETHER

ORGANIZATIONAL STRUCTURES, MEASURES AND EXCHANGE FORMATS

Enabling collaborations

- **Tools for continuous exchange and collaboration**
(collaboration software)
- **Regular in-person meetings and workshops**
(e.g. annual workshops using the Dagstuhl seminar format)
- **Exchange program**
(e.g. exchange stays of ten days at individual institutions)
- **Cluster Conference and Hackathon** each Cycle
- International exchange and collaboration



DIFFERENT FROM BUT ALSO COMPLEMENTARY TO THE NFDI



NFDI Nationale Forschungsdaten Infrastruktur (National Research Data Infrastructure)

- **Mission:** Data from science and research are systematically accessed, networked and made usable in a sustainable and qualitative manner for the entire German science system
- *KonsortSWD* aims to expand and integrate the existing infrastructure for research data management to improve Findability, Accessibility, Interoperability, and Reuse of social science data.
- *BERD@NFDI* aims to build a powerful platform for collecting, processing, analyzing and preserving business, economic and related data & will facilitate the integrated management of algorithms and data along the whole research cycle, with a special focus on unstructured (big) data such as video, image, audio, text or mobile data.

InfPP: New Data Spaces for the Social Sciences

- **Mission:** Unlocking new types of data for the social sciences, developing new methods, technologies, and procedures for collecting and analyzing social science data and conducting research on the methodological quality of these new data types

Complementarity of InfPP and KonsortSWD/NFDI

- **InfPP establishes** innovative ways of generating data that can be **fitted into NFDI** structures of findable, accessible, interoperable, and reusable data.
- An **interface** ensures the exchange of innovative developments within the InfPP and the entire NFDI.

NÄCHSTE SCHRITTE

2023

2024

Call for Proposals (IfW)

14. März

Entscheidung/Bewilligung

März

Go-live der Website

(www.new-data-spaces.de)

14. März

Start des Programms

(Start des Zentralprojekts voraussichtlich im April)

Juni

Ideenskizzen für
Workshop 12. Mai

17. April

Workshop

(Einladung am 24.04.)

12. Mai

Einreichen von Anträgen

20. September

Begutachtung der Anträge

ab Dezember

NEUEN DATENRÄUME FÜR DIE SOZIALWISSENSCHAFTEN

DFG Infrastrukturschwerpunktprogramm

- **DFG:**
 - Eckard Kämper
- **Programme Committee**
 - Cordula Artelt, LfBi Bamberg (Speaker)
 - Stefan Liebig, FU Berlin
 - Corinna Kleinert, LfBi Bamberg
 - Alexander Mehler, GU Frankfurt
 - Reinhard Pollak, GESIS Mannheim



KSWD, 27.03.2023 | 17:45 – 18:15 Uhr

www.new-data-spaces.de