

Vorgaben des FDZ der BA im IAB zur Durchführung von Datenfernverarbeitung und Gastaufenthalten

Das Wichtigste auf einen Blick

Auswertungsprogramme (Do-Files)

- Müssen nach den Vorgaben des FDZ erstellt werden (siehe Kapitel 2)
- Do-Files dürfen keine Daten enthalten
 - Kopieren Sie niemals Ergebnisse, Betriebs- oder Personennummern in Ihre Do-Files
 - Externe Daten müssen als Datensatz ans FDZ geschickt werden (siehe Kapitel 3.2)

Aufträge zur Datenfernverarbeitung mit JoSuA (Job Submission Application)

- 1 Auswertungsjob pro Tag
- Maximal 20 Do-Files pro Auftrag
- Warten Sie vor dem Einschicken eines neuen Auftrags bitte unbedingt, bis der vorherige Job in Ihrem Projekt freigegeben ist und/oder Ihr Gastaufenthalt beendet ist.
- Auswahl des Modus in JoSuA:
 - Internal Use*
 - Für alle vorbereitenden Schritte der Datenaufbereitung und -auswertung
 - Ergebnisse dürfen nicht kopiert, abgeschrieben oder veröffentlicht werden
 - Ergebnisse dürfen nicht von Personen eingesehen werden, die nicht in der Datennutzungsvereinbarung des Projekts genannt sind.
 - Presentation / Publication*
 - Dient der Veröffentlichung von Ergebnissen für Vorträge oder Publikationen
 - Zur Übertragung von Do-Files vor und nach Gastaufenthalt

Gastaufenthalt

- Die über JoSuA hochgeladenen Programm- und Ergebnisdateien liegen auf einem eigenen Server und werden nicht automatisch in die Projektverzeichnissen im Gästernetz kopiert.
- Do-Files müssen mindestens 3 Arbeitstage vor dem ersten Aufenthaltstag über JoSuA im Modus *Presentation / Publication* hochgeladen werden.

Datennutzung

- Der verfügbare Arbeitsspeicher ist auf **30 GB pro Nutzer und Projekt** beschränkt. Achten Sie auf Ihren Arbeitsspeicher, um Ihren und die Auswertungsjobs anderer Nutzerinnen und Nutzer nicht zu beeinträchtigen.
- Die Bereitstellung externer aggregierter Daten in Ihrem Projektverzeichnis kann nach erfolgreicher Prüfung bis zu 3 Arbeitstage dauern.

Ergebnisse

- Die Datenschutzprüfung für Auswertungen aus dem Modus *Presentation / Publication* kann bis zu 5 Arbeitstage in Anspruch nehmen.
- Ergebnisse vorheriger *Internal-Use-Jobs* sind nicht mehr verfügbar, sobald ein Job im Modus *Presentation / Publication* abgeschlossen wurde.
- Falls Sie aggregierte Datensätze als Ergebnisdatei erhalten möchten, sprechen Sie die Vorgehensweise bitte **vorher** mit uns ab.

Bei Fragen kontaktieren Sie bitte das Forschungsdatenzentrum (iab.fdz@iab.de). Bitte geben Sie immer Ihre FDZ-Projektnummer (fdzXXXX) und ggf. die Job-ID aus JoSuA an.

Inhalt

1	Datenzugang am FDZ	1
1.1	Projektverzeichnisse auf den FDZ-Servern	1
1.2	JoSuA (Job Submission Application)	1
2	Gestaltung der Programmdateien	2
2.1	Die Master-Datei	2
2.2	Auswertungsprogramme	3
2.3	Dokumentation der Fallzahlen in Auswertungen	4
2.4	Dokumentation von Aggregationsschritten	5
2.5	Datensparsamkeit	5
3	Weitere Ressourcen	7
3.1	Ado-Files	7
3.2	Zuspielen externer aggregierter Datensätze	7
4	Ablauf der Datennutzung	8
4.1	Testen der Programme	8
4.2	Datenfernverarbeitung	8
4.3	Auswahl des Nutzungsmodus in JoSuA	8
4.4	Vorbereitungen für den Gastaufenthalt	9
4.5	Rücksendung der geprüften Ergebnisse nach Gastaufenthalt	10
5	Ergebnisübermittlung	11
5.1	Datenschutzrichtlinien	11
5.2	Zulässige Ausgabeformate für das Log-File	11
5.3	Grafiken	12
5.4	Übermittlung von generierten aggregierten Datensätzen	12
6	Publikation der Ergebnisse	13
7	Sonderregelungen der Softwarenutzung	14
8	Bedienungsanleitung JoSuA	16
8.1	Login	16
8.2	Start	16
8.3	My Jobs	16
8.4	Projekte	18
8.5	Hochladen von Aufträgen	21

1 Datenzugang am FDZ

1.1 Projektverzeichnisse auf den FDZ-Servern

- Jedes Forschungsprojekt erhält **pro** FDZ-Datensatz ein eigenes Verzeichnis (fdzXXXX). Verschiedene FDZ-Datensätze können nicht gemeinsam verwendet werden.
- Jedes Projektverzeichnis enthält folgende Unterverzeichnisse:

orig: In diesem Verzeichnis werden alle beantragten Originaldaten bereitgestellt. Auch ggf. eingesendete Aggregatdaten (siehe Abschnitt 3.2) befinden sich hier. Für dieses Verzeichnis erhalten Sie nur Leserechte.

data: Hier werden alle generierten Datensätze abgelegt. Pro Verzeichnis sind **maximal 30 GB** erlaubt. Sollte das Maximum überschritten werden, können keine weiteren Jobs eingeschickt werden, bis der Speicherbedarf reduziert wurde.

prog: In diesem Ordner befinden sich alle Auswertungsprogramme für den Gastaufenthalt (Do-Files) sowie ggf. weitere Dateien (z. B. ado-Dateien).

log: Dieser Ordner enthält alle Ergebnisdateien des Gastaufenthalts, inklusive Grafiken.

doc: Dieser Ordner ist für eingeschickte Dokumente im PDF-Format und zur Erstellung von Notizen während des Gastaufenthalts.

- Auf die Datenverzeichnisse orig und data können Programme auch über JoSuA zugreifen. Für Do- und Log-Files gibt es in JoSuA einen separaten Speicherort. Do- und Log-Files werden nicht automatisch zwischen JoSuA und dem Projektverzeichnis abgeglichen.
- Das Anlegen weiterer Verzeichnisse oder Unterverzeichnisse ist nicht erlaubt.

1.2 JoSuA (Job Submission Application)

- Das FDZ nutzt die webbasierte Software JoSuA (Job Submission Application) des Instituts zur Zukunft der Arbeit (IZA) für
 - die Datenfernverarbeitung,
 - das Hochladen von Do-Files vor Gastaufenthalt und
 - die Übermittlung von Ergebnissen und Do-Files nach Gastaufenthalt.
- Jeder Nutzer bzw. jede Nutzerin erhält nach Abschluss eines Nutzungsvertrags eine Kennung für JoSuA. Die Weitergabe des Passworts an Dritte ist nicht erlaubt.
- In JoSuA kann auf Daten in den Verzeichnissen **orig** und **data** zurückgegriffen werden. Beachten Sie hierzu die Programmierhilfe in Beispiel 2.
- Do- und Log-Files werden von JoSuA nicht mit den Inhalten der Verzeichnisse prog und log gespiegelt, sondern in einer separaten Ablage gespeichert.
- Die Bedienung von JoSuA wird detailliert in Kapitel 8 beschrieben.

2 Gestaltung der Programmdateien

2.1 Die Master-Datei

- Alle im Verlauf des Projektes verwendeten Programme sind von einer übergeordneten Master-Datei aufzurufen. Sie muss immer **master.do** heißen. Beispiel 1 veranschaulicht ihren Aufbau.
- Benutzen Sie zum Zugriff auf Ihre Dateien immer die Stata-Globals **\$orig**, **\$prog**, **\$data** und **\$log**. Sie stehen sowohl bei der Datenfernverarbeitung mit JoSuA als auch bei Gast-aufenthalten automatisch zur Verfügung und verweisen auf den korrekten Speicherort. Die Globals dürfen nicht überschrieben werden.
- Nutzen Sie keine cd-Befehle oder relative Pfadangaben.
- Verwenden Sie **niemals** die Anweisung **set max_memory**.
- Nicht erlaubte Befehle werden durch JoSuA zensiert. In diesen Fällen sehen Sie die Zeile auskommentiert mit ***** CENSORED *****.
- Die Master-Datei muss alle im Projektverlauf relevanten Auswertungsprogramme enthalten. Es ist wichtig, dass in der Master-Datei hinter jedem Programmaufruf eine kurze Beschreibung der Inhalte des jeweiligen Unterprogramms erfolgt.
- Die Auswertungsprogramme sind in der Reihenfolge des Programmablaufs zu nummerieren. Arbeiten mehrere Nutzer/-innen getrennt an einem Projekt, so können ihre Programme durch Kürzel im Dateinamen unterschieden werden, z.B. *th01_PersDat.do*.
- Alle von Ihnen erzeugten Dateien (Ergebnisdateien, Analysedatensätze, usw.) müssen anhand Ihrer Programme und der Originaldaten replizierbar sein.
- Bereits ausgeführte Programmschritte, die seit der letzten Einsendung nicht verändert wurden, sind im Projektverlauf auszukommentieren. Über JoSuA besteht der Zugriff auf gespeicherte Datensätze in den Verzeichnissen orig und data. Lediglich Do- und Log-Files werden auf separaten Servern gespeichert.
- Für JoSuA empfehlen wir, im Modus *Internal Use* die Zeilenlänge mit dem Befehl *set linesize* auf 120 zu setzen, da dies der Breite des Output-Fensters in JoSuA entspricht.

Beispiel 1: Inhalt der Datei master.do

```
* Master-Datei des Projektes "Projektname"
* Nutzer: T. Huber (th) und W. Mueller (wm)

version 13
clear
set more off // damit das Programm durchlaeuft
*set linesize 120 // Max.Zeilenlaenge Internal Use
*set linesize 255 // Max.Zeilenlaenge Presentation/Publication

*do "$prog/th01_PersDat.do" // Aufbereitung Personendaten
*do "$prog/th02_BetrDat.do" // Aufbereitung Betriebsdaten
*do "$prog/th03_DatXY.do" // Zusammenspielen
*do "$prog/th04_AnalyseA.do" // Untersuchung Hypothese 1 (...)
*do "$prog/th05_AnalyseB.do" // Untersuchung Hypothese 2 (...)
*do "$prog/wm01_Datagen.do" // Aufbereitung Personendaten
do "$prog/wm02_AnalyseX.do" // Untersuchung Hypothese 5 (...)
do "$prog/wm03_AnalyseY.do" // Untersuchung Hypothese 6 (...)

* Ausgabe einer Dateiliste in eine log-Datei
capture log close
log using "$log/filelist.log", replace
dir $prog\*
dir $log\*
dir $data\*
dir $orig\*
log close
```

2.2 Auswertungsprogramme

- Ergebnisdateien (auch Grafiken) müssen den gleichen Dateinamen tragen wie das Programm, mit dem diese Dateien generiert werden. Beispielsweise erzeugt **01_PersDat.do** die Ergebnisdatei **01_PersDat.log**.
- Die Programme müssen für die Datenschutzprüfung dokumentiert sein (siehe Beispiel 2). Achten Sie dabei darauf,
 - Zwischenüberschriften zu verwenden (z. B. "Aufbereitung der Variablen X", "Aggregation auf Betriebsebene"),
 - die Inhalte und Funktion von Schleifen zu beschreiben,
 - für alle wichtigen Variablen, die neu gebildet wurden, Variablenlabels und ggf. Wertelabels zu vergeben,
 - möglichst sprechende Variablennamen zu verwenden und
 - selbst generierte Variablen direkt vor Auszählungen kurz zu beschreiben.

Beispiel 2: Inhalt eines Programms zur Datenauswertung

```
* Logfile schreiben:
capture log close
log using "$log/th04_AnalyseA.log", replace

* Programm-Datei des Projektes "Projektname" - th04_AnalyseA.do
* Inhalt: Untersuchung Hypothese 1

* Aggregationsniveau: Branchen nach Bundesland
*
* aggregiert wird in th03_DatXY.do
* mind. 20 Betriebe pro Datenzeile
* die Personenzahl pro Zeile steht in Faelle

clear all

* Ausgangsdaten einlesen:
use "$orig/Datensatzname", clear

* Analysen
[...]

* Erzeugte Daten speichern:
save "$data/neuename", replace
log close
```

2.3 Dokumentation der Fallzahlen in Auswertungen

- Im Rahmen der Datenschutzprüfung am FDZ (siehe Kapitel 5.1) werden sämtliche Werte, die auf weniger als 20 Beobachtungen (Personen und/oder Betrieben) basieren, gelöscht. Um das Rückrechnen von gelöschten Werten zu verhindern, werden weitere Werte gelöscht oder gerundet.
- Daher sind bei allen statistischen Kennzahlen oder multivariaten Analysen die zugrunde liegenden Fallzahlen auszuweisen. Bei Fehlen der Fallzahlen werden die betreffenden Ergebnisse gelöscht.
- Bei der Ausgabe von Mittelwerten müssen neben der Fallzahl immer auch Minimum, Maximum und Standardabweichung ausgegeben werden.
- Werden Quoten dargestellt, muss die Anzahl der gültigen Werte in den einzelnen Subgruppen ebenfalls ausgewiesen werden.
- Bei Verwendung von Hochrechnungs- oder Gewichtungsfaktoren in deskriptiven Analysen sind auch immer analoge ungewichtete Ergebnisse auszuweisen. Die gewichteten und die dazugehörigen ungewichteten Ergebnistabellen müssen dabei immer direkt untereinander stehen, weil dies die Datenschutzprüfung erleichtert und somit beschleunigt. Bitte berücksichtigen Sie dies auch in Schleifen.
- Bei Jobs im Modus *Presentation / Publication* ist vor jeder multivariaten Auswertung eine Deskription zu den Variablen, die ins Modell eingehen, erforderlich.
- Bei Auswertungen kombinierter Betriebs- und Personendaten muss in deskriptiven Tabellen neben der Personenzahl auch die Anzahl der Betriebe ausgewiesen werden.

2.4 Dokumentation von Aggregationsschritten

- Werden die ursprünglichen Mikrodaten aggregiert (z.B. auf die Ebene von Betrieben, Regionen etc.), muss dies im Analyseprogramm entsprechend beschrieben werden.
- Sollen diese Aggregate später dargestellt werden, müssen die zugrundeliegenden Fallzahlen ebenfalls berechnet werden:
 - Es muss **je aggregierter Variable** eine zusätzliche Variable gebildet werden, welche die Anzahl der entsprechenden gültigen Werte (Personen und Betriebe) im Aggregat enthält (z.B. **bysort Aggregat: egen N_var1 = count(var1)**).
 - Ist das Aggregat eine Quote, muss die Anzahl der gültigen Werte in den einzelnen Subgruppen des Aggregats berechnet werden (z.B. Anzahl Männer in Bundesland X und Anzahl Frauen in Bundesland X für die Frauenquote in Bundesland X).
- Alle im Anschluss an eine Aggregation erstellten Ergebnisdateien müssen am Anfang folgende Informationen enthalten:
 - auf welche Ebene aggregiert wurde,
 - in welchem Programmschritt aggregiert wurde (bitte auch in der Master-Datei kennzeichnen),
 - wie viele Personen und Betriebe mindestens pro Datenzeile eingegangen sind (z.B.: Zellen mit <20 Betrieben wurden gelöscht) und
 - Name der Variablen, welche die Beobachtungsanzahl pro Datenzeile enthält.
- Fehlen die entsprechenden Hinweise auf die Aggregation zu Beginn der Ergebnisdatei, wird bei der Datenschutzprüfung eventuell zu viel gelöscht.
- Die Anzahl der aggregierten Tabellen sollte, im Sinne der Datensparsamkeit, auf ein Minimum beschränkt sein.

2.5 Datensparsamkeit

- Verwenden Sie den Speicherplatz in Ihrem Verzeichnis **data** sparsam. Wird das Maximum von **30 GB** überschritten, können keine weiteren Jobs eingeschickt werden, bis der Speicherbedarf reduziert wurde.
- Speichern Sie nur dann Zwischendatensätze ab, wenn es für die weitere Analyse erforderlich ist. Datensätze sollten mithilfe Ihrer Do-Files jederzeit reproduzierbar sein. Löschen Sie zwischenzeitlich generierte Datensätze, wenn diese nicht länger benötigt werden (siehe Beispiel 3).
- Schränken Sie Ihren Datensatz auf diejenigen Variablen und Beobachtungen ein, die Sie für Ihre Analysen tatsächlich benötigen (siehe Beispiel 4 und Beispiel 5).
- Zur Entwicklung Ihrer Programme kann es sinnvoll sein, die Lauffähigkeit zunächst an einer Stichprobe zu testen (z. B. mit dem Befehl **sample**).
- Nutzen Sie den Befehl **compress** beim Speichern von Datensätzen.

- Der verfügbare Arbeitsspeicher ist beschränkt auf **30 GB pro Nutzer und Projekt**.

Beispiel 3: Löschmodul

```
* Programm-Datei zum Loeschen nicht mehr benoetigter Dateien

capture log close
log using "$log/07_delete.log", replace

* Dateien loeschen:
capture erase "$data\test1.dta" // erase "$data\test*" funktioniert NICHT
capture erase "$data\test2.dta"
[etc.]

log close
```

Beispiel 4: Panel aus BHP mit Jahr, Betriebsnummer und Anzahl Beschäftigte

```
use betnr az_ges using $orig\bhp_7514_m06_p50_v1_1975.dta, clear
gen int jahr = 1975 // Festlegung des Datentyps integer statt float
forvalues j = 1976/2014 {
    append using $orig\bhp_7514_m06_p50_v1_`j'.dta, keep(betnr az_ges)
    replace jahr = `j' if missing(jahr)
}
```

Beispiel 5: Verknüpfen der Beschäftigten aus SIAB mit Basisbetriebsdatei

```
* Beschaeftigtenspells auswaehlen:
use if quelle == 1 using $orig\siab_7514_v1.dta, clear

keep persnr betnr begepi tentgelt // Auswahl interessierender Variablen
keep if year(begepi) >= 2000 // Einschraenken auf Spells ab 2000

* Variable w08_3 aus Basisbetriebsdatei zuspielen:
gen int jahr = year(begepi)
merge m:1 betnr jahr ///
    using $orig\siab_7514_v1_bhp_basis_v1.dta, keepusing(w08_3)
keep if _merge ==3
drop _merge
```

3 Weitere Ressourcen

3.1 Ado-Files

- Ado-files können nicht online heruntergeladen werden. Das FDZ stellt ausgewählte Pakete lokal zur Verfügung (siehe Liste unter http://doku.iab.de/fdz/access/stata_ado.pdf).
- Verwenden Sie zur Installation von am FDZ verfügbaren Paketen folgenden Befehl, der alle zugehörigen Dateien in Ihr prog-Verzeichnis kopiert: ¹

```
net install packagename
```

- Wenn Sie weitere Ado-Files oder eigene Ado-Files verwenden möchten, schicken Sie diese per E-Mail an das FDZ-Postfach.
- Ado-Files werden im Ordner prog gespeichert. Die Ado-Files im Ordner prog stehen auch in JoSuA zur Verfügung. Das Verzeichnis ist sowohl beim Gastaufenthalt als auch in JoSuA automatisch als *adopath* eingebunden. Ado-Files im prog-Verzeichnis werden damit automatisch erkannt. Eine Verwendung des Befehls *adopath* ist nicht notwendig.
- Wenn Sie Ado-Files ausschließlich im Rahmen der Datenfernverarbeitung verwenden möchten, laden Sie diese in JoSuA hoch. Ado-Files werden im Reiter *Projects* (siehe Abschnitt 8.4) einmalig hochgeladen. Sie stehen dann für die Datenfernverarbeitung zur Verfügung, nicht jedoch im Gastaufenthalt.

3.2 Zuspielen externer aggregierter Datensätze

- Externe Kenngrößen auf aggregiertem Niveau (z.B. Arbeitslosenquoten nach Kreisen) dürfen den Daten zugespielt werden, wenn sie den Datenschutzrichtlinien des FDZ entsprechen.
- Jeder aggregierte Wert muss auf mindestens 3 Beobachtungen beruhen. Das Zusammenspielen von Einzelbeobachtungen ist nicht möglich. Der Datensatz muss Variablen enthalten, die die Anzahl der zugrundeliegenden Beobachtungen angeben.
- Die entsprechenden Kenngrößen **müssen** dem FDZ mit einer Beschreibung des Datensatzes (inkl. der Merkmale, der Aggregationsebene und der Quellenangabe) **als Datensatz** übermittelt werden (iab.fdz@iab.de). Die Übermittlung aggregierter Kennzahlen innerhalb von Do-Files ist nicht zulässig.

¹ Derzeit nur im Gastaufenthalt und nicht über JoSuA möglich.

4 Ablauf der Datennutzung

4.1 Testen der Programme

- Überprüfen Sie Ihre Auswertungsprogramme vor Gastaufenthalt und Datenfernverarbeitung mit Hilfe der auf der FDZ-Homepage zur Verfügung gestellten Testdaten.
- In JoSuA und im Gastaufenthalt sind die Globals für Pfadangaben automatisch definiert. Für Ihre eigene Testumgebung müssen Sie die Globals **\$orig**, **\$data**, **\$prog**, **\$log** und ggf. den **adopath** definieren, bevor Sie die Datei master.do starten. Schreiben Sie diese Global-Definitionen **nicht** in master.do, sondern in eine Datei **profile.do**. Diese wird von Stata automatisch ausgeführt und muss nicht durch master.do angesprochen werden. Sie kann beispielsweise im aktuellen Arbeitsverzeichnis abgelegt werden (siehe auch <http://www.stata.com/help.cgi?profile>). Die Datei darf **nicht** in JoSuA hochgeladen werden.
- Führen Sie die Datei master.do in Ihrer Testumgebung aus.
- Nach erfolgreichem Test laden Sie Ihre Programme **unverändert** für die Datenfernverarbeitung in JoSuA hoch.

4.2 Datenfernverarbeitung

- Die von Ihnen in JoSuA hochgeladenen Auswertungsprogramme müssen **vollständig lauffähig** sein. Das FDZ führt keine Fehlerbehebungen durch.
- Ihre Auswertungsprogramme werden mit der aktuellsten am FDZ vorhandenen Stata-Version gestartet.
- Bitte warten Sie mit dem Hochladen der nächsten Auswertung, bis die Ergebnisse des letzten vorangegangenen Jobs innerhalb Ihres Projekts bereitgestellt wurden.
- Die Rücksendung ist abhängig von Serverauslastung, Rechen- und Prüfaufwand. Bitte beachten Sie dies bei Ihrer Arbeitsplanung.

4.3 Auswahl des Nutzungsmodus in JoSuA

- Grundsätzlich stehen in JoSuA zwei Nutzungsmodi zur Verfügung:
 - 1) Internal Use
 - Die Ergebnisse sind nur für den „internen“ Gebrauch, d. h. sie dienen nur zur Entwicklung der Auswertungsprogramme.
 - Hochgeladene Aufträge werden automatisch gestartet und die Ergebnisse per Bildschirmanzeige verfügbar gemacht.
 - Es erfolgt eine skriptbasierte Datenschutzprüfung.
 - Ergebnisse dürfen nicht kopiert, ausgedruckt oder anderweitig entnommen werden.

- Ausgeschlossen ist außerdem eine Weitergabe der Informationen an Personen, die nicht in der Datennutzungsvereinbarung des Projekts genannt sind.
- Die Nutzung der automatisch erzeugten Ergebnisse im Modus *Internal Use* in Veröffentlichungen stellt einen Verstoß gegen die Datenschutzvereinbarung dar.
- Bitte verwenden Sie den Modus *Internal Use* solange, bis Sie Ihre Endergebnisse oder Zwischenergebnisse für Präsentationen oder Publikationen erstellt haben.

2) Presentation/Publication

- In diesem Modus werden herunterladbare Log-Dateien erstellt.
 - Der Modus ist auszuwählen, wenn Ergebnisse im Rahmen einer Präsentation oder Publikation veröffentlicht werden sollen.
 - Die hochgeladenen Do-Files werden von den Mitarbeiterinnen und Mitarbeitern des FDZ gestartet und manuell auf die Einhaltung des Datenschutzes geprüft.
 - Laden Sie nur die für Präsentationen oder Publikationen notwendigen Programme hoch. Kommentieren Sie Befehle aus, deren Ergebnisse nicht benötigt werden, oder stellen Sie den Befehlen das Kommando "quietly" voran.
- Um eine Rückrechnung und Ergänzung von Ergebnissen auszuschließen, sind die Ergebnisse vorheriger Aufträge im Modus *Internal Use* nicht mehr verfügbar, sobald ein Auftrag im Modus *Presentation / Publication* eingeschickt wird.
 - Bitte beachten Sie, dass zur Gewährleistung des Datenschutzes im Modus *Internal Use* einige Befehle grundsätzlich gesperrt werden.² Ausnahmeregelungen sind nicht möglich. Sie haben in diesen Fällen die Möglichkeit, durch Auswahl des Modus *Presentation / Publication* manuell geprüfte Ergebnisse zu erhalten.

4.4 Vorbereitungen für den Gastaufenthalt

- Wenn Sie während Ihres Gastaufenthalts am FDZ vorbereitete Auswertungsprogramme nutzen wollen, laden Sie diese **spätestens drei Arbeitstage** vor dem ersten Aufenthaltstag über JoSuA im Modus *Presentation / Publication* hoch. Bitte schreiben Sie in das Kommentarfeld, dass nur Do-Files für den Gastaufenthalt hochgeladen werden.
- Fügen Sie in der Datei master.do vor dem Aufruf weiterer Do-Files die Zeile „**STOP**“ ein oder kommentieren Sie den Aufruf weiterer Do-Files aus. Dadurch wird verhindert, dass der Job Rechenzeit benötigt und Ergebnisse produziert. Der Job muss gestartet werden, um die hochgeladenen Programme in JoSuA verfügbar zu machen.
- Fehlen die beschriebenen Anpassungen in der Datei master.do, laufen automatisch alle Programme durch, die dort aufgerufen werden. Dies kann, je nach Laufzeit der Programme,

² Derzeit ist der Befehl list im Modus *Internal Use* gesperrt. Das FDZ behält sich die Sperrung weiterer Befehle vor.

zu Verzögerungen bei der Bereitstellung der Do-Files führen, da die Prüfung und Freigabe immer erst nach Beendigung des Programmlaufs erfolgen kann.

- Alle hochgeladenen Do-Files können nach erfolgter Eingangsprüfung durch das FDZ während des Gastaufenthalts von einem internen Server heruntergeladen werden.
- Sie finden im Gastaufenthalt einen Link zu JoSuA auf dem Desktop des Stata-Servers.
- Aggregatdatensätze sowie ado-Files, die Sie für Ihren Gastaufenthalt benötigen, müssen ebenfalls **spätestens drei Arbeitstage** vor dem ersten Aufenthaltstag zur Prüfung an das FDZ gesendet werden. Nennen Sie bitte das Datum des geplanten Gastaufenthalts in der E-Mail. Falls Sie innerhalb von zwei Arbeitstagen keine Eingangsbestätigung erhalten haben, geben Sie uns bitte Bescheid. Beachten Sie die Abschnitte zu ado-Files (Abschnitt 3.1) und externen Aggregatdaten (Abschnitt 3.2).

4.5 Rücksendung der geprüften Ergebnisse nach Gastaufenthalt

- Nach dem Gastaufenthalt müssen Sie Ihre Programme nochmals mit JoSuA starten. Die im Verlauf des Gastaufenthalts erzeugten Ergebnisse werden nicht herausgegeben.
- Laden Sie am Ende Ihres Gastaufenthaltes alle während des Gastaufenthalts bearbeiteten Do-Files im Modus *Presentation / Publication* in JoSuA hoch.
- Sorgen Sie dafür, dass master.do und alle Programmdateien vollständig lauffähig sind.
- Sie finden im Gastaufenthalt einen Link zu JoSuA auf dem Desktop des Stata-Servers.
- Nach einer manuellen Datenschutzkontrolle erhalten Sie Zugang zu Ihren Ergebnissen und allen hochgeladenen Do-Files.
- Lassen Sie alle Do-Files vollständig laufen, wenn Sie die Ergebnisse sofort für Publikationen oder Präsentationen benötigen. Sind Sie nur an den Do-Files interessiert und möchten nicht, dass das Programm durchläuft, dann schreiben Sie in master.do ein „STOP“ vor den Aufruf der Do-Files oder kommentieren Sie den Aufruf der Do-Files aus. Alle hochgeladenen Jobs werden automatisch durch das FDZ gestartet.
- Alle hochgeladenen Do-Files stehen nach erfolgter Datenschutzprüfung in JoSuA zum Download zur Verfügung. Eine nachträgliche Übermittlung von Do-Files durch das FDZ ist nicht möglich.

5 Ergebnisübermittlung

5.1 Datenschutzrichtlinien

- Die erzeugten Ergebnisse werden vom FDZ auf die Vorgaben zum Datenschutz geprüft und danach an die Nutzerinnen und Nutzer übermittelt.
- Beachten Sie die Hinweise zur Dokumentation von Fallzahlen in Abschnitt 2.3 sowie von Aggregationsschritten in Abschnitt 2.4.
- Alle Ergebnisse, die den Datenschutzvorgaben des FDZ nicht entsprechen, werden gelöscht.
- Detaillierte Angaben zu den Richtlinien und Beispiele sind im [FDZ-Methodenreport \(06/2012\)](#) von Hochfellner et al. (2012)³ aufgeführt.
- Der Nutzer bzw. die Nutzerin verpflichtet sich, keine Rückrechnungen vorzunehmen, um gelöschte Werte zu ermitteln.
- Bei Verstößen wird der Nutzer bzw. die Nutzerin für die Dauer von bis zu zwei Jahren von jeder weiteren Datennutzung ausgeschlossen. Des Weiteren können die im Nutzungsvertrag genannten Strafen zur Anwendung kommen.
- Informationen über Verstöße gegen die o. g. Verpflichtungen werden an andere Forschungsdatenzentren weitergegeben.

5.2 Zulässige Ausgabeformate für das Log-File

- Stata-Ergebnisdateien müssen entweder die Endung „.log“ oder „.txt“ tragen.
- Ergebnisse, die in separate Dateien außerhalb der von Stata angelegten Log-Files geschrieben wurden, können nicht bereitgestellt werden. Sie müssen wieder in das Log-File eingebunden werden (siehe Beispiel 6). Zur Datenschutzprüfung müssen die Ergebnisse unmittelbar vor dem eingebundenen Text nochmals in lesbarer Form dargestellt werden.

³ Hochfellner, Daniela; Müller, Dana; Schmucker, Alexandra; Roß, Elisabeth (2012): Datenschutz am Forschungsdatenzentrum, FDZ-Methodenreport 06/2012 (de).

Beispiel 6: Einbindung von Ergebnissen in das Log-File

```
* Inhalt einer csv- oder txt-Datei im Log-File ausgeben:  
[...] // Ausgabe als Tabelle/Liste  
type $log/ergebnisse.csv  
type $log/ergebnisse.txt  
  
* LaTeX-Code direkt ins Log-File, nicht in separate Datei schreiben:  
[...] // Ausgabe der Regressionen  
esttab A B, c(b se), tex // Beispiel für Regressionen A und B
```

5.3 Grafiken

- Stata-Grafiken (.gph) können nur bereitgestellt werden, wenn sie mit der Option **asis** erzeugt wurden (siehe Beispiel 7). Die Formate wmf, eps, und png sind ebenfalls zulässig.
- Im Modus *Internal Use* kann nur das Grafikformat **png** dargestellt werden.
- Bitte generieren Sie nur dann Grafiken, wenn es nicht möglich ist, die Grafiken durch die Ergebnisdateien z.B. in Excel selbst zu erstellen. Da der Aufwand der Datenschutzprüfung von Grafiken sehr hoch ist, sollte die Anzahl der erstellten Grafiken so gering wie möglich gehalten werden.
- Bei jeder Grafik ist nachzuweisen, auf wie vielen Beobachtungen die einzelnen dargestellten Werte basieren. Dies kann über die Angabe der Fallzahlen entweder direkt in den Grafiken oder in Tabellen, die sich unmittelbar vor oder nach der Erstellung der Grafiken befinden, erfolgen.
- Speichern Sie die Grafiken im Verzeichnis **log** ab.

Beispiel 7: Ausgabe von Grafiken

```
* Ausgabe von Grafiken im zulaessigen Format  
graph ... , ... saving($log\mygraph, replace asis)  
* oder  
graph save $log\mygraph, replace asis
```

5.4 Übermittlung von generierten aggregierten Datensätzen

- Falls Sie aggregierte Datensätze als Ergebnisdatei erhalten möchten, sprechen Sie die Vorgehensweise bitte **vorher** mit uns ab.
- Die Generierung der aggregierten Daten sollte während eines Gastaufenthalts erfolgen. Die Programme müssen anschließend nochmals mit JoSuA gestartet werden. Bitte verwenden Sie dafür den Modus *Presentation / Publication* und vermerken Sie im Kommentarfeld, dass Sie einen generierten Datensatz erhalten möchten und den entsprechenden Dateinamen.
- Für die Dokumentation der aggregierten Datensätze gelten die Vorgaben für aggregierte Variablen entsprechend.
- Für jedes Projekt kann **nur einmal** ein solcher Datensatz weitergegeben werden. Dies gilt auch, wenn der aggregierte Datensatz in einem anderen FDZ-Projektverzeichnis zur Zupielung bereitgestellt werden soll. Die Übermittlung des Datensatzes erfolgt per E-Mail.

6 Publikation der Ergebnisse

- In Publikationen sind Daten und Dokumentationen entsprechend den Vorgaben auf der Homepage des FDZ (<http://doku.iab.de/fdz/access/Zitierweisen.pdf>) zu zitieren.
- Senden Sie bitte ein Belegexemplar an das FDZ.

7 Sonderregelungen der Softwarenutzung

- Standardmäßig bietet das FDZ für seine Nutzerinnen und Nutzer im Gastaufenthalt und in der Datenfernverarbeitung nur Stata als Analysesoftware an. In Ausnahmefällen kann andere Software vom FDZ zur Verfügung gestellt werden. Setzen Sie sich dazu bitte frühzeitig mit dem FDZ in Verbindung. Alle Aufträge, die z. B. auf Matlab oder R basieren, müssen per E-Mail eingendet werden. Bitte nutzen Sie diesen Modus des Datenfernzugriffs nur für diejenigen Teile Ihrer Analysen, die nicht in Stata ausgeführt werden können.
- Bei Verwendung anderer Software als Stata gilt grundsätzlich, dass alle Datenaufbereitungs- und Analyseschritte, die mit Stata machbar sind, auch mit Stata durchgeführt werden müssen. Des Weiteren müssen die Programme für die Sondersoftware aus einem Stata Do-File aufgerufen und gestartet werden.
- Da die Nutzung von Sondersoftware über JoSuA nicht möglich ist, müssen Programmdateien für die **Datenfernverarbeitung** oder **vor Gastaufenthalt** per E-Mail an iab.fdz@iab.de verschickt werden. Hierbei müssen folgende Regelungen beachtet werden:
 - Zur Bearbeitung Ihrer Aufträge benötigen wir in jeder E-Mail Angaben zum Projekt in der Betreffzeile (siehe Beispiel 8).
 - Neu eingeschickte Dateien überschreiben bereits vorhandene, wenn sie den gleichen Dateinamen haben.
 - Unter *Sonstiges* bitte schon in der Betreffzeile auf Dinge hinweisen, die bei der Bearbeitung der Programme zu beachten sind (z.B. „Matlab“).
 - Wir bitten um Verständnis dafür, dass wir Aufträge unbearbeitet zurücksenden, wenn relevante Informationen in der Betreffzeile fehlen oder die Formvorgaben nicht erfüllt sind.
 - Senden Sie für ein Projekt keine neuen Programme ein, solange Sie die Ergebnisse der letzten Datenfernverarbeitung noch nicht erhalten haben. Nutzen Sie in dieser Zeit auch nicht JoSuA, um weitere Ergebnisse im Modus *Presentation / Publication* zu erzeugen.
- Nach einem Gastaufenthalt:
 - Nach einem Gastaufenthalt werden nicht die während des Aufenthalts erzeugten Ergebnisse herausgegeben, sondern die entsprechenden Programme werden vom FDZ-Team nochmals gestartet. Sorgen Sie daher dafür, dass master.do und die anderen Programmdateien vollständig lauffähig sind.
 - Wenn Sie nach Abschluss Ihres Gastaufenthalts Programmdateien und/oder Ergebnisse erhalten wollen, schicken Sie eine E-Mail an das FDZ-Postfach (siehe

Beispiel 8). Schreiben Sie zusätzlich „nach GA“ in die Betreffzeile und geben Sie in der E-Mail das Datum des Gastaufenthaltes sowie die Programm-Dateien, die wir Ihnen schicken sollen, an.

- Wir starten master.do und Sie erhalten die geprüften Ergebnisse, die in diesem Programmdurchlauf erzeugt wurden. Außerdem erhalten Sie standardmäßig die Programm-Dateien, die Sie während Ihres Aufenthalts erstellt oder verändert haben.

Beispiel 8: Betreffzeile der E-Mail bei Datenfernverarbeitung mit Sondersoftware

Projektnummer – max. Speicherbedarf – erwartete Laufzeit – Datum Vertragsende – Sonstiges

fdz###, # GB, # Stunden/Tage, yyyy-mm-dd, [Sonstiges]

yyyy-mm-dd = Datum im Format Jahr-Monat-Tag

Beispiele:

fdz041, 3 gb, 14 h, 2016-01-31, Matlab

fdz041, 3 gb, 14 h, 2016-01-31, nach GA am 2014-04-30, Matlab

8 Bedienungsanleitung JoSuA

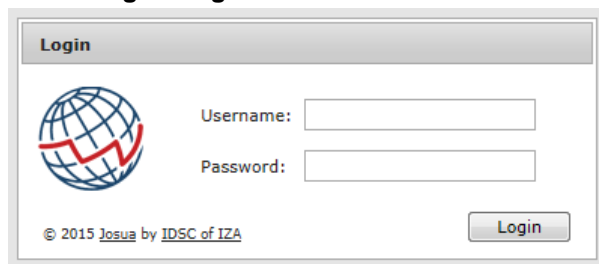
Die folgenden Abschnitte beschreiben, wie man durch die JoSuA-Webschnittstelle navigiert, Aufträge einsendet und anschließend auf die Ergebnisse zugreift.

Aktuelle Antworten auf häufige Fragen zu JoSuA finden Sie auch auf der FDZ-Homepage unter http://fdz.iab.de/de/FDZ_Data_Access/FDZ_Remote_Data_Access/faq.aspx.

8.1 Login

- Rufen Sie die Internetseite www.josua.iab.de/gui/index.html auf.
- Geben Sie im Login-Fenster (siehe Abbildung 1) die Zugangsinformationen ein, die Sie vom FDZ erhalten haben.
- Ihr Benutzername und Initialpasswort sind identisch mit den Anmeldeinformationen, die Sie bei Gastaufenthalten nutzen.
- Falls Sie noch keine Zugangsdaten erhalten haben oder diese vergessen haben, setzen Sie sich mit dem FDZ in Verbindung (iab.fdz@iab.de).
- Die Weitergabe des Passworts an Dritte ist nicht erlaubt.⁴

Abbildung 1: Login



8.2 Start

Nach der erfolgreichen Anmeldung erscheint die Nutzeroberfläche. Das Hauptmenü besteht aus mehreren Reitern: *My Jobs*, *Projects* und *Job Submission*. Diese werden in den folgenden Abschnitten näher beschrieben. Zu Beginn ist der Reiter *My Jobs* ausgewählt. In der oberen rechten Ecke können Sie Ihr Passwort ändern, die Hilfe aufrufen oder sich ausloggen.

8.3 My Jobs

Der Reiter *My Jobs* (Meine Aufträge), der in Abbildung 2 dargestellt ist, zeigt eine Liste mit allen aktuellen Aufträgen und deren Status, z. B. *queue* (Warteschlange), *running* (Prozess läuft) oder *released* (abgeschlossen). Klicken Sie auf die Lupe, um Details des ausgewählten Auftrags anzusehen.

⁴ Bei mehreren Nutzenden pro Projekt erhält jeder Nutzende eigene Zugangsdaten.

Abbildung 2: Reiter *My Jobs* (Meine Aufträge)

Sent in	Project	Status	Censoring	Options
2016-06-10 12:42:05	FDZ9999	success	released	
2015-12-15 16:42:21	FDZ9999	failed	released	
2015-12-09 12:51:42 (presentation)	FDZ9999	success	released	

Showing 1 to 3 of 3 entries (filtered from 50 total entries)

Abbildung 3: Detailansicht zum ausgewählten Auftrag

→ Job Info

Owner John Doe	Submitted By web	
Status success	Censoring released	
Sent In 2016-06-10 12:42:05	Processing Started 2016-06-10 12:42:07	Processing Ended 2016-06-10 12:42:14

Input Files
Input files are only visible when running jobs in 'Presentation / Publication' mode.

Output Files

In der Detailansicht (siehe Abbildung 3) finden Sie unter *Input Files* alle hochgeladenen Dateien und unter *Output Files* alle verfügbaren Ergebnisdateien. Die Darstellung ist abhängig vom Einsendemodus. Zudem sind einige allgemeine Daten zu Ihrem Auftrag aufgeführt.

- **Input Files (Hochgeladene Dateien):**

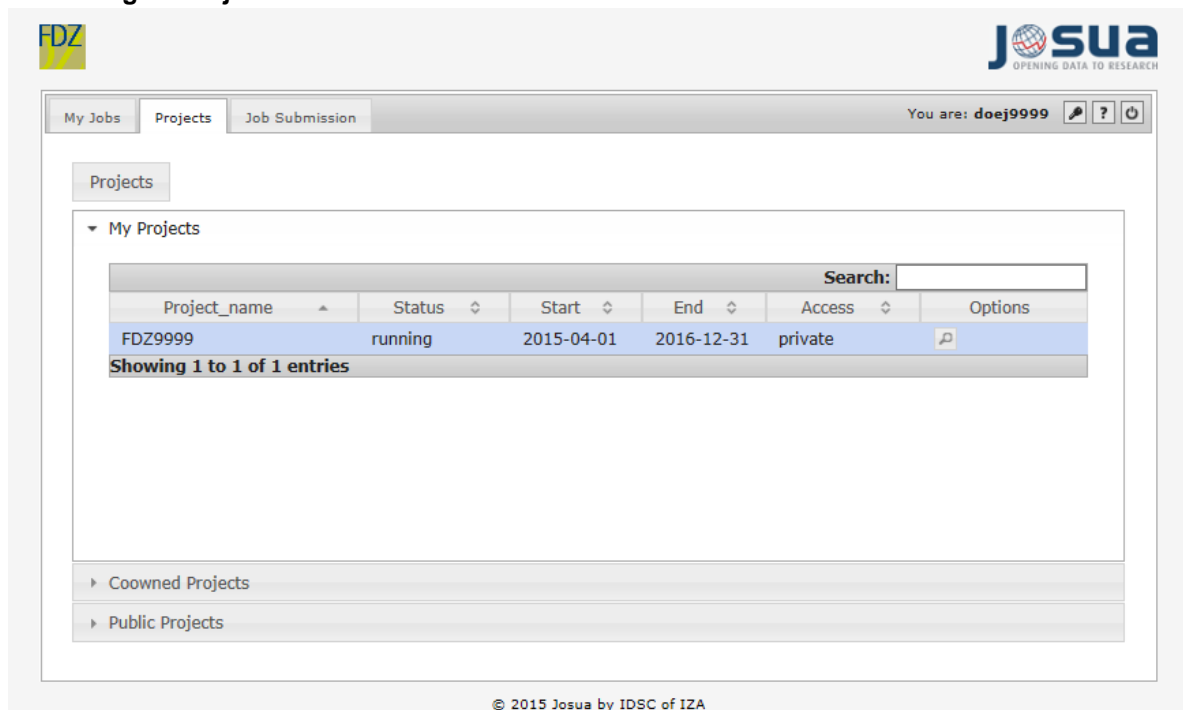
- *Internal Use:* In diesem Modus ist die Download-Funktion grundsätzlich deaktiviert.

- *Presentation / Publication*: Alle hochgeladenen Do-Files für diesen Auftrag werden aufgelistet. Dateien können per Klick auf den Dateinamen oder *Download* heruntergeladen werden. Der Link erscheint nach Abschluss der Datenschutzprüfung.
- **Output Files (Ergebnisdateien)**:
 - *Internal Use*: Sie können die Ergebnisse Ihres Auftrags per Klick auf eines der Vorschaubilder einsehen. Navigieren Sie mittels Klicken der Richtungspfeile durch die Vorschaufenster.
 - *Presentation / Publication*: Nach der manuellen Datenschutzprüfung können alle geprüften Ergebnisse als Textdatei heruntergeladen werden. Der Link erscheint erst nach erfolgter Datenschutzprüfung. Um Dateien herunterzuladen klicken Sie auf den Dateinamen oder auf *Download* und speichern die Datei auf Ihrem Computer.

8.4 Projekte

Im Reiter *Projects* (Projekte) können Sie Eckdaten Ihres Auftrages einsehen, Aufträge anderer Projektmitglieder (sofern vorhanden) anzeigen und Ado-Files hochladen. Ein Klick auf *My Projects* (Meine Projekte) oder *Co-owned Projects* (Gemeinschaftsprojekte)⁵ zeigt Ihnen das Projekt an, bei dem Sie aktuell angemeldet sind. Öffentliche Projekte gibt es derzeit keine. Eine beispielhafte Ansicht ist in Abbildung 4 dargestellt.

Abbildung 4: Projektansicht



The screenshot shows the 'Projects' view in the Josua system. At the top, there are navigation tabs for 'My Jobs', 'Projects', and 'Job Submission'. The user is logged in as 'doej9999'. The main content area is titled 'Projects' and contains a section for 'My Projects'. Below this, there is a search bar and a table with the following data:

Project_name	Status	Start	End	Access	Options
FDZ9999	running	2015-04-01	2016-12-31	private	[icon]

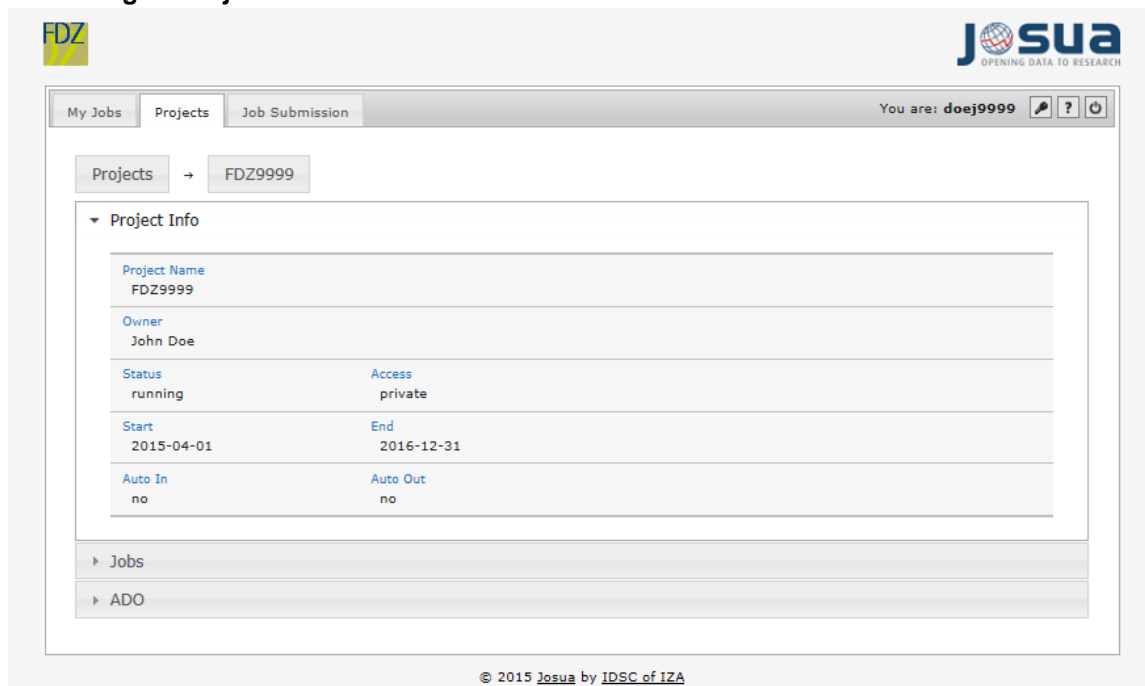
Below the table, it says 'Showing 1 to 1 of 1 entries'. There are also sections for 'Coowned Projects' and 'Public Projects' which are currently empty.

⁵ My Projects und Co-owned Projects sind im Wesentlichen dasselbe, nur dass Sie in ersterem Fall als Erstbesitzer und im zweiten Fall als Mitbesitzer eingetragen sind. Beide haben die gleichen Rechte.

Durch Klick auf die Lupe in der Spalte Optionen gelangen Sie zu den Projektdetails. In diesem Menü finden Sie drei Dropdown-Listen:

- **Project Info (Projektinformationen):** Zeigt alle verfügbaren Informationen des ausgewählten Projektes an, wie in Abbildung 5 dargestellt ist.
- **Jobs (Aufträge):** Listet die Aufträge aller Nutzerinnen und Nutzer des gewählten Projektes auf. Zusätzlich zu generellen Informationen können Sie durch Anklicken der Lupe dieselben Optionen aufrufen, die Sie auch unter *My Jobs* finden.
- **ADO:** Laden Sie Ado-Files für Ihr Projekt hoch (siehe Abbildung 6). Jeder Nutzer kann Dateien hinzufügen, die alle autorisierten Mitglieder für ihre Aufträge nutzen können. Falls dort schon von anderen Mitgliedern hochgeladene Dateien sind, können Sie diese hier sehen. Klicken Sie auf den Mülleimer, um Dateien zu löschen. JoSuA wird Sie auffordern, die Löschung zu bestätigen. Um neue Dateien hinzuzufügen klicken Sie auf *Add file (Datei hinzufügen)* und suchen Sie die Datei auf Ihrer Festplatte oder geben Sie den korrekten Pfad an. Klicken Sie anschließend auf *Upload (Hochladen)*, um Ihre Auswahl zu speichern. Falls bereits eine Datei mit gleichem Namen existiert, wird diese überschrieben.

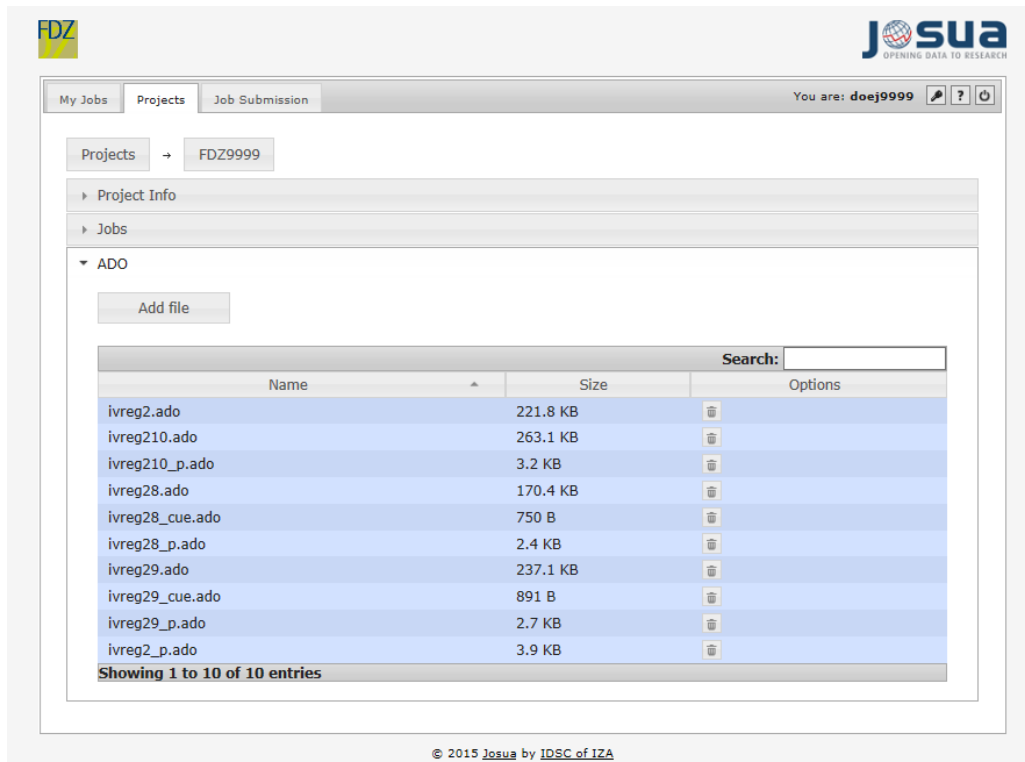
Abbildung 5: Projektinformationen



The screenshot shows the 'Project Info' section of the Josua interface. The project name is 'FDZ9999' and the owner is 'John Doe'. The status is 'running' and the access is 'private'. The start date is '2015-04-01' and the end date is '2016-12-31'. The 'Auto In' and 'Auto Out' options are both set to 'no'. Below the project info, there are expandable sections for 'Jobs' and 'ADO'.

Project Name	FDZ9999		
Owner	John Doe		
Status	running	Access	private
Start	2015-04-01	End	2016-12-31
Auto In	no	Auto Out	no

Abbildung 6: Ado-Dateien



The screenshot shows the Josua web interface. At the top, there are navigation tabs: "My Jobs", "Projects", and "Job Submission". The user is logged in as "doej9999". The current view is "Projects" for project "FDZ9999". Under "Jobs", the "ADO" section is expanded, showing a list of files. A search bar is present above the table. The table lists 10 .ado files with their names, sizes, and options (trash icon). Below the table, it says "Showing 1 to 10 of 10 entries".

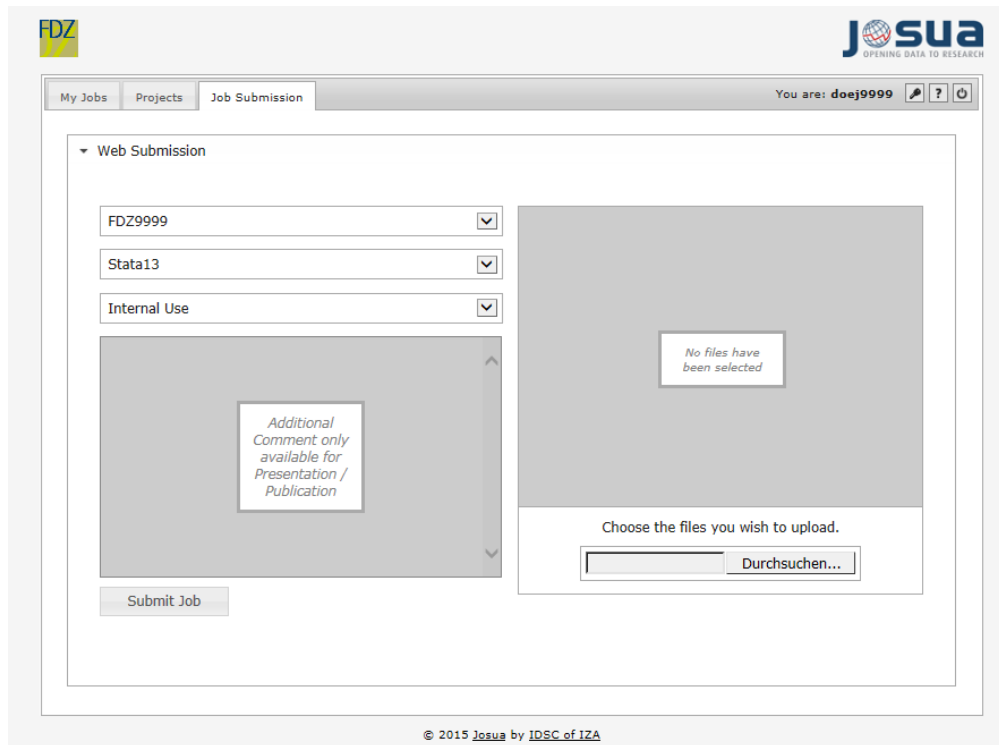
Name	Size	Options
ivreg2.ado	221.8 KB	
ivreg210.ado	263.1 KB	
ivreg210_p.ado	3.2 KB	
ivreg28.ado	170.4 KB	
ivreg28_cue.ado	750 B	
ivreg28_p.ado	2.4 KB	
ivreg29.ado	237.1 KB	
ivreg29_cue.ado	891 B	
ivreg29_p.ado	2.7 KB	
ivreg2_p.ado	3.9 KB	

Showing 1 to 10 of 10 entries

© 2015 Josua by IDSC of IZA

8.5 Hochladen von Aufträgen

Abbildung 7: Hochladen von Aufträgen



The screenshot shows the 'Job Submission' page in the JoSuA interface. At the top, there are navigation tabs for 'My Jobs', 'Projects', and 'Job Submission'. The user is logged in as 'doej9999'. The main content area is titled 'Web Submission' and contains several form elements: three dropdown menus for project ID (FDZ9999), software package (Stata13), and submission mode (Internal Use). Below these is a text area for an 'Additional Comment only available for Presentation / Publication'. A 'Submit Job' button is located at the bottom left. On the right side, there is a large grey area with the message 'No files have been selected' and a file selection interface with a 'Durchsuchen...' button. The footer contains the copyright notice '© 2015 JoSuA by IDSC of IZA'.

Ein Auftrag zur Datenfernverarbeitung wird übermittelt, indem alle auszuführenden **Do-Files** im Reiter *Job Submission* (Hochladen von Aufträgen) hochgeladen werden. Eine Master-Datei namens **master.do** muss enthalten sein.

Auf der linken Seite ist das Projekt aufgeführt, an dem Sie gegenwärtig arbeiten (z. B. fdz9999) und das Software-Paket, das genutzt wird. Im dritten Drop-Down-Menü im linken Ausschnitt können Sie auswählen, ob der Auftrag im Modus *Internal Use* oder *Presentation / Publication* eingeschickt werden soll (siehe Abschnitt 4.3). Wenn Sie letzteren Modus auswählen, können Sie im darunter liegenden Feld einen Kommentar eingeben, der relevante Details oder Erklärungen enthält. Verzichten Sie auf Umlaute und Sonderzeichen und verwenden Sie maximal 512 Zeichen.

Klicken Sie anschließend auf der rechten Seite unten auf *Durchsuchen...*, um Do-Files von Ihrer Festplatte in JoSuA hochzuladen. Die hochgeladenen Dateien werden aufgelistet. Sie können die Liste prüfen und einzelne Dateien entfernen. Wenn Sie mit dem Hochladen Ihrer Do-Files fertig sind, klicken Sie auf *Submit Job* (Auftrag übermitteln), um Ihren Auftrag zu übersenden.