UNIONE EUROPEA
FONDO EUROPEO DI SVILUPPO REGIONALE

Ministero Sviluppo Economico

REGIONE PUGLIA
DIPARTIMENTO SVILUPPO ECON
INNOVAZIONE, ISTRUZION
FORMAZIONE E LAVORO

TECHNICAL REPORT 4.1

PROGETTAZIONE DELL'ESPERIMENTO

INDICE

INDICE 3

1. STORIA DEL DOCUMENTO 5

2. INTRODUZIONE 6

3. RICERCA EMPIRICA 8

3.1 INTRODUZIONE 8
3.2 RICOGNIZIONE 8
3.3 CASO DI STUDIO 10
3.4 CONFRONTO TRA INDAGINI SPERIMENTALI 13
3.5 DEFINIZIONI PRELIMINARI 15

4. LINEE GUIDA PER L'ESECUZIONE DI UNA RICERCA EMPIRICA 17

4.1 ESPERIMENTO CONTROLLATO 17
4.2 DEFINIZIONE 18
4.3 CONTESTO SPERIMENTALE 19
4.4 PIANIFICAZIONE 20
4.5 IPOTESI SPERIMENTALI 21
4.6 SELEZIONI DELLE VARIABILI 23
4.7 SELEZIONI DEI SOGGETTI SPERIMENTALI 24
4.8 DESIGN SPERIMENTALE 25
4.9 PROGETTO DELL'INDAGINE 32
4.10 STRUMENTI 37
4.11 ESECUZIONE 37
4.12 ANALISI ED INTERPRETAZIONE 38
4.13 STATISTICHE DESCRITTIVE 40
4.14 TEST DELLE IPOTESI 42
4.15 PRESENTAZIONE E CONFEZIONAMENTO DEI RISULTATI 45

5. CASO DI STUDIO 47

5.1 INTRODUZIONE AL CASO DI STUDIO 47
5.1.1 Caratteristiche delle Metodologie di Ricerca 49

5.1.2 Processo per eseguire un caso di studio *51*
5.1.3 Terminologia *52*
5.2 PROGETTAZIONE E PIANIFICAZIONE DI UN CASO DI STUDIO 53
5.2.1 Definire il caso *53*
5.2.2 Protocollo del Caso di Studio *54*
5.2.3 Checklist *55*
5.3 RACCOLTA DEI DATI 56
5.3.1 Fonti per la raccolta dei dati *56*
5.3.2 Checklist *57*
5.4 ANALISI DEI DATI 57
5.4.1 Analisi Quantitativa dei Dati *57*
5.4.2 Analisi Qualitativa dei dati *58*
5.4.3 Validità dei risultati *59*
5.4.4 Checklist *60*
5.5 REPORT DEI RISULTATI 60
5.5.1 Caratteristiche *61*
5.5.2 Struttura *62*
5.5.3 Checklist *65*

6. RICERCA AZIONE 66

6.1 ACCORDO TRA RICERCATORI E COMMITTENTE (RCA) 68
6.2 MODELLO DI PROCESSO CICLICO 68
6.3 TEORIA 70
6.4 CAMBIARE ATTRAVERSO L'AZIONE 71
6.5 APPRENDERE ATTRAVERSO LE RIFLESSIONI 71

7. LE INDAGINI DEL PROGETTO I-MOLE 73

7.1 MANUTENZIONE RAPIDA ED AFFIDABILE 73
7.2 COMPOSIZIONE DEI SERVIZI 74
7.3 SPECIFICAZIONE DEI SERVIZI 75
7.4 RICERCA DI UN SERVIZIO 76
7.5 USABILITÀ ED UX DEL PORTALE 77
7.6 PENETRATION TEST 80
7.7 SICUREZZA CON LA BLOCKCHAIN 82

8. BIBLIOGRAFIA 85

1. STORIA DEL DOCUMENTO

DATA	AUTORE	VERSIONE	CONTENUTI & MODIFICHE

2. INTRODUZIONE

Le indagini empiriche mirano a sperimentare l'efficacia delle innovazioni introdotte in Assioma.

Esse sono condotte con l'approccio **costruttivista**. L'approccio consente di produrre **teorie locali** che sono valide per Assioma e che possono essere modificate nel tempo man mano che i progetti che si eseguono nell'azienda, anche se non sono sperimentali, producono nuova conoscenza. Il tipo di indagine preminente è il **caso di studio** essenzialmente per due ragioni: la sperimentazione deve avvenire nel contesto reale; spesso, i risultati della sperimentazione richiedono l'esecuzione di attività incluse nei progetti sperimentali ma non implicate nella sperimentazione.

Esse sono state eseguite sempre con i **Minimum Viable Product** (**MVP**) per evitare sprechi nel produrre prototipi operativi (prodotto da sperimentare) che evidenziano carenze di efficacia e quindi devono essere modificati. In altri termini, si spende il minimo possibile per arrivare ad un prototipo che possa essere verificato, testato o sperimentato così che se il risultato della validazione dovesse essere negativo si è speso poco e, comunque, si è compreso prima se e cosa debba essere modificato.

Quando è possibile, le indagini sono condotte con il paradigma della **Ricerca-Azione**. La comunità di destinatari dell'innovazione riconosce come "proprio e valido" il problema per cui si sta sviluppando l'innovazione e si fa promotrice delle ipotesi di soluzione. Così che se dopo la sperimentazione i risultati non sono valutati, dalla stessa comunità, soddisfacenti si decidono le iniziative di miglioramento con un processo di progettazione partecipata. I ricercatori di Assioma promuovo e facilitano il processo mentre i veri attori dell'innovazione e delle indagini correlate sono gli stakeholders.

I cantieri sperimentali aperti hanno caratteristiche diverse, dipendentemente dagli obiettivi dell'innovazione da sperimentare. Pertanto, ogni cantiere sperimentale ha un suo processo di sperimentazione che è adeguato agli obiettivi dell'innovazione.

Nel seguito di questo report si descrivono, per completezza, il paradigma di *ricerca-azione* e il *caso di studio*. Dopo la descrizione di questi concetti si descrivono i cantieri sperimentali attivati e per ognuno di essi si riporta il progetto dell'indagine sperimentale.

Il rapporto è strutturato secondo le seguenti sezioni:

- **Ricerca empirica**. Analizza le diverse strategie di indagine empirica con le relative caratteristiche per scegliere la strategia più adeguata ad un contesto sperimentale.
- **Linee guida per l'esecuzione di una ricerca empirica**. Il processo più adeguato a una indagine empirica in un ambiente industriale come i-MOLE.
- **Caso di Studio** e **Ricerca Azione**. sono due sezioni che dettagliano le due strategie relative ai titoli. Questo dettaglio è necessario perché esse sono le strategie più utilizzate nelle indagini sperimentali di i-mole.
- **Bibliografia**. Riferimenti bibliografici utilizzati per la elicitazione dei concetti espressi nel rapporto. Quando non c'è corrispondenza biunivoca tra concetto e riferimento bibliografico, quest'ultimo non è indicato nel testo ma è comunque citato nella bibliografia per dare al lettore la possibilità di approfondire quanto riportato nel rapporto.

3. RICERCA EMPIRICA

3.1 Introduzione

In letteratura si identificano tre tipi di ricerca empirica:

- Ricognizione (Survey)
- Caso di Studio (Case Study)
- Esperimento controllato (Experiment)

Le caratteristiche fondamentali delle indagini sono di seguito rappresentate

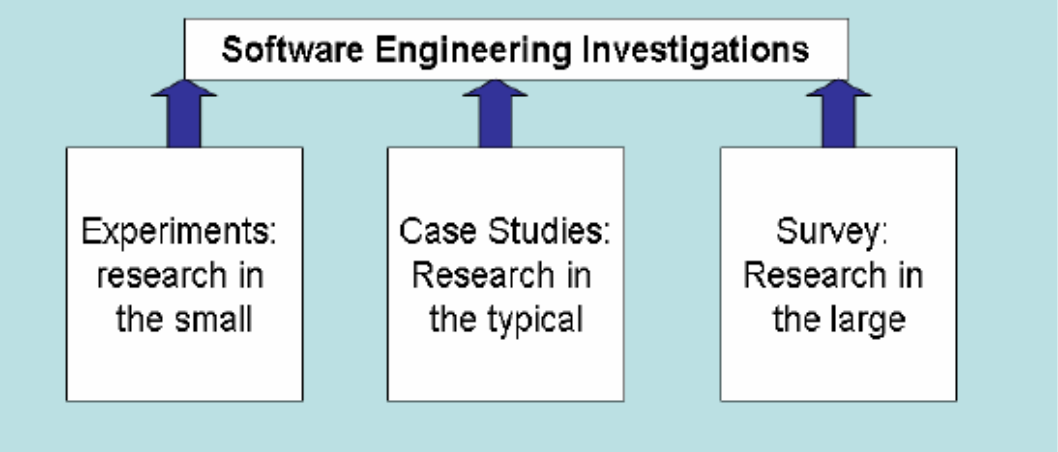

Figura 3.1 – Tipi di ricerca empirica

Nei prossimi paragrafi si descrivono le caratteristiche di ciascun tipo di ricerca.

3.2 Ricognizione

Una ricognizione può consistere:

- nell'analisi di dati raccolti in progetti svolti
- nell'analisi di opinioni di esperti rilevate attraverso questionari o interviste
- in analisi che utilizzano sia dati che opinioni dei partecipanti ai progetti, in cui sono stati raccolti i dati; in quest'ultimo caso le

interviste hanno lo scopo di comprendere meglio, dagli esperti, gli eventi accaduti e le iniziative realizzate durante l'esecuzione del progetto.

I dati utilizzati nella ricognizione, in genere, sono stati raccolti durante l'esecuzione dei progetti con obiettivi diversi da quelli della ricerca empirica. Questo è un punto di forza a favore dei risultati della ricerca; infatti, questi si basano su dati che certamente non hanno subito l'influenza del ricercatore e degli obiettivi della ricerca. Inoltre, la raccolta dei dati avviene per motivi inerenti all'amministrazione dei progetti in cui sono stati rilevati; pertanto, le ricognizioni risultano a bassi costi. Quando si decide di eseguire una ricognizione su progetti eseguiti, i dati sono già tutti disponibili quindi la ricerca non presenta rischi per essa stessa, per le unità e per i soggetti sperimentali. A questi vantaggi si contrappone la limitatezza dell'evidenza empirica che è possibile ricavare da questo tipo di ricerca. Infatti, nell'esecuzione delle ricognizioni non si ha controllo sul contesto e questo rende impossibile la replica dell'indagine e difficile l'uso dei risultati per generalizzare le tesi validate con altre indagini. Ad esempio: un'organizzazione ha usato un linguaggio Object Oriented per la prima volta; al termine del progetto una ricognizione retrospettiva, basata sui dati raccolti durante l'esecuzione del progetto, può ricavare la relazione tra dimensione del codice prodotto, l'effort speso, il numero di fault e failure; questi risultati possono essere confrontati con quelli di altri progetti che usavano linguaggi procedurali per verificare se l'approccio Object Oriented ha generato miglioramenti; qualunque sia il risultato che si ottiene questo non è affidabile perché i progetti sono difficilmente paragonabili visto che molte variabili dei progetti non sono state controllate né sono determinabili con l'accuratezza necessaria.

Le indagini che utilizzano i dati rilevati in progetti eseguiti sono, spesso, chiamate indagini post-mortem. Esse possono raccogliere informazioni relative a grandi progetti oppure ad un gran numero di progetti, perciò, spesso si chiamano ricerca in grande. La ricerca può avere due tipi di obiettivi:

- *Descrittivo*: per formulare asserzioni caratterizzanti su una popolazione. Es. per determinare la distribuzione di caratteristiche o attributi.

- *Esplicativo*: mira a spiegare i comportamenti della popolazione. Es. cercare di spiegare perché, sulla base di una serie di variabili esaminate, alcuni sviluppatori preferiscono una tecnica piuttosto che un'altra.

3.3 Caso di studio

I casi di studio sono *esperimenti in vivo*, ovvero, sono eseguiti su servizi reali. Pertanto, la ricerca attraverso il caso di studio si deve adattare alla realtà produttiva osservata. Più precisamente, i valori che le variabili di processo possono assumere sono dipendenti dai processi in esecuzione nelle filiere produttive dell'impresa. Di conseguenza, le variabili non possono essere completamente definite dal ricercatore ma sono fortemente vincolate dal ciclo produttivo dell'impresa che mette a disposizione il campo sperimentale. Ovvero, le variabili indipendenti non sono definite o modificate dalla decisione del ricercatore ma dalla decisione del responsabile del progetto o dall'evoluzione naturale delle variabili di processo durante l'esecuzione del progetto osservato. E', altresì, evidente che la ricerca attraverso il caso di studio deve essere progettata in modo che si possa compiere in una finestra temporale determinata dalle esigenze del progetto reale in cui si indaga.

I casi di studio sono usati con due prospettive:

- *osservare* l'efficacia di una tecnologia impiegata e le cause che generano i valori rilevati degli indicatori di efficacia;
- *comparare* l'efficacia di due tecnologie concorrenti; in genere, l'attuale e quella innovativa e le cause che generano la differenza di efficacia.

Nella prima prospettiva la tecnologia impiegata può essere:

- l'innovativa, allora il caso di studio mira a rafforzare l'evidenza empirica della capacità di miglioramento dei servizi o dei processi che conferiscono i risultati di ricerca e la relativa tecnologia sperimentata e se non ci riuscisse sarebbe sconsigliata l'istituzionalizzazione della tecnologia;

- l'attuale, allora il caso di studio mira a constatare che essa non raggiunge l'efficacia attesa; se ciò risultasse empiricamente verificato la ricerca servirebbe a rilevare le necessità di miglioramento della stessa tecnologia altrimenti essa servirebbe a confermare la validità della tecnologia e ad assicurare che non c'è bisogno di innovazione.

Nella seconda prospettiva il caso di studio serve a stabilire la migliore tra le due tecnologie paragonate.

Per analizzare i risultati di un caso di studio è necessario disporre di termini di riferimento con i quali confrontare gli indicatori di efficacia che si ricavano durante le indagini.

Ci sono tre modi per ricavare la baseline:

- con i dati raccolti in progetti eseguiti dall'impresa. A questo scopo potrebbe essere utilizzata una ricognizione su dati già disponibili. La baseline così ricavata contiene i termini di riferimento cercati. Ad esempio: si deve sperimentare una tecnologia che ha lo scopo di migliorare la qualità dei progetti software che è valutata con un determinato set di misure. Prima dell'indagine si analizzano i progetti software già presenti nel patrimonio dell'impresa e si rilevano le misure che appartengono al set definito; da queste si ricavano i valori di base che l'impresa normalmente raggiunge. Nel caso di studio devono essere rilevate le stesse misure con valori migliori di quelli della baseline perché la tecnologia sperimentata possa essere considerata efficace.
- Con un sister-project il servizio coinvolto nel caso di studio usa il nuovo metodo paragonandolo con un metodo già radicato nell'impresa. Entrambi i servizi debbono avere le medesime

caratteristiche, in modo tale da essere confrontabili tra loro. Ad esempio: si deve sperimentare una nuova tecnologia che migliora la manutenzione dei sistemi software; si eseguono le stesse manutenzioni sugli stessi sistemi software in un progetto che usa la nuova tecnologia e con quello sister che usa la tecnologia consueta per l'impresa, avendo cura che gli sviluppatori impegnati nei due progetti abbiano capacità analoghe; i dati derivati da questo secondo progetto costituiscono la baseline a cui paragonare i risultati rilevati con il primo progetto.

- Con i segmenti di progetto o di prodotto, se la tecnologia è applicabile per blocchi, si scelgono a caso i segmenti in cui è applicata la nuova tecnologia e quelli in cui si applica la tecnologia consueta per l'impresa; i dati rilevati con la seconda classe di segmenti costituiscono la baseline. Ad esempio: si deve sperimentare una nuova tecnologia per la progettazione del software onde migliorare la qualità del progetto misurata con un set di misure; il sistema è diviso in sottosistemi alcuni dei quali, scelti a caso, sono progettati con le tecnologie consuete e gli altri con quelle innovate; le misure rilevate dal primo blocco di sottosistemi costituisce la baseline.

Nell'esecuzione di un caso di studio, è necessario minimizzare i fattori che possono falsare la valutazione dei risultati. Questi non sono controllati dalla tecnologia che si sta sperimentando ma possono generare risultati che sono accreditati ai fattori controllati dalla tecnologia sperimentata. In questo caso i risultati dell'esperimento sono, almeno, parzialmente falsati. Tali fattori collaterali costituiscono quelle che i ricercatori chiamano *minacce dell'esperimento*. Per esempio, se si volesse misurare l'efficacia di un nuovo tool, il progetto dell'esperimento dovrebbe essere tale da consentire la distinzione tra quanta influenza ha il tool sull'efficacia e quanta ne ha l'esperienza degli sviluppatori; infatti, l'esperienza degli sviluppatori

consentirebbe di produrre progetti con indicatori di qualità di gran lunga migliori di quelli causati dall'uso del tool.

Un caso di studio può essere eseguito su un *servizio pilota* o può essere eseguito su un *servizio esplorativo*. È opportuno usare il servizio pilota quando l'innovazione ha già un buon livello di maturità, i ricercatori la trasferiscono agli sviluppatori professionisti che la applicano nel servizio pilota. Durante l'esecuzione di questo servizio sono misurate le variabili che esprimono l'efficacia dell'innovazione nei processi produttivi dell'impresa che sperimenta.

Si consiglia il servizio esplorativo, invece, quando la tecnologia non è sufficientemente matura. Pertanto, durante l'erogazione del servizio sperimentale essa è migliorata con il contributo del team che partecipa alla sperimentazione come soggetti sperimentali. Al termine di questo esperimento, l'opportunità di diffondere l'innovazione è valutata sulla base sia della maturità raggiunta dalla tecnologia, sia dell'efficacia che la stessa ha mostrato nel progetto sperimentale. Dalle precedenti considerazioni si evince che il caso di studio è costoso e rischioso. Infatti, si devono affiancare al progetto reale attività complementari che servono solo per l'indagine e che quindi aumentano il costo del progetto. Inoltre, l'esecuzione di tali attività non garantisce il risultato atteso e, quindi, il loro prodotto potrebbe costituire uno scarto di produzione. Questo oltre ad aumentare il costo aumenta il rischio di non terminare il progetto reale nei tempi previsti.

Infine, per la carenza di controllo delle variabili indipendenti della ricerca sussiste anche il rischio, per il ricercatore, di non ricavare dati quantitativamente e qualitativamente sufficienti per provare le ipotesi che costituiscono lo scopo della ricerca.

3.4 Confronto tra indagini sperimentali

La Tabella 3.1 confronta le indagini sperimentali rispetto ad un insieme di diversi fattori:

Fattore	Ricognizione	Caso di Studio	Esperimenti
Controllo sulla esecuzione	Nessuno	Basso	Alto

Controllo sulle misure	Nessuno	Alto	Alto
Costi della ricerca	Basso	Medio	Alto
Facilità di replica	Alto	Basso	Alto

Tabella 3.1 – Confronto tra indagini

Controllo sulla esecuzione: il controllo che il ricercatore ha sullo studio.

Ad esempio, in un case study i dati sono raccolti durante la esecuzione di un progetto. Se il progetto è interrotto (per motivi economici o decisioni di management) il ricercatore non può continuare. Nell'esperimento il ricercatore ha il pieno controllo sulla esecuzione.

Controllo sulle misure: il potere decisionale che il ricercatore ha sulle misure da raccogliere. Nella survey si possono raccogliere solo misure sulla opinione del campione, nel case study bisogna attenersi alla realtà tipica dell'impresa, nell'esperimento il ricercatore ha minori restrizioni di scelta.

Costi della ricerca: dipende dalla dimensione dello studio e dalle risorse necessarie. La survey richiede poche risorse. Il case study sono eseguite su un progetto in esecuzione (on-line investigation) con possibile profitto. L'esperimento ha come risultato esperienza e conoscenza acquisita (off-line investigation), senza ritorni immediati di investimenti. Facilità di replica: mira a validare i risultati dello studio originario per poter generalizzare i risultati ad una più ampia popolazione. L'esperimento consiste nel replicare lo studio con le stesse caratteristiche su una popolazione diversa, per confermare la confidenza dei risultati del primo studio.

Ciascuna ricerca empirica è svolta in un contesto specifico. La figura 3.2 illustra i diversi contesti di esecuzione di una ricerca empirica:

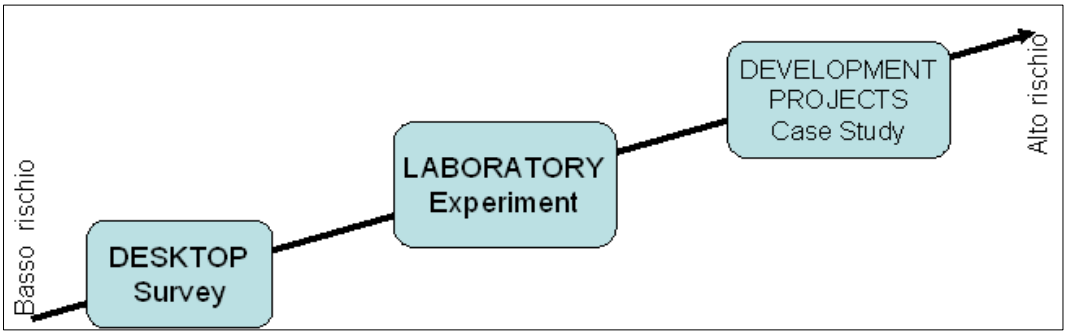

Figura 3.2 – Contesto di esecuzione di una ricerca empirica

Si individuano complessivamente, tre contesti di applicazione:

- *Desktop*: l'innovazione è valutata off-line senza eseguire il processo modificato. Valutazioni di questo tipo non coinvolgono soggetti che devono applicare il metodo, tool ecc. È appropriato eseguire una survey attraverso interviste e analisi retrospettive della letteratura.
- *Laboratorio*: l'innovazione è valutata in un contesto controllato (off-line). Il processo è eseguito e controllato.
- *Development Projects*: l'innovazione è valutata su un progetto reale on-line, es. progetti pilota.

3.5 Definizioni preliminari

Prima di illustrare il processo per la pianificazione di una ricerca, è necessario dare alcune definizioni di termini che saranno utilizzati nell'ambito del documento.

In un esperimento si desidera studiare il risultato ottenuto dopo aver variato il valore delle variabili di input al processo.

La figura 3.3 rappresenta la relazione tra gli elementi coinvolti in una ricerca sperimentale. Per completezza, si fornisce una breve definizione.

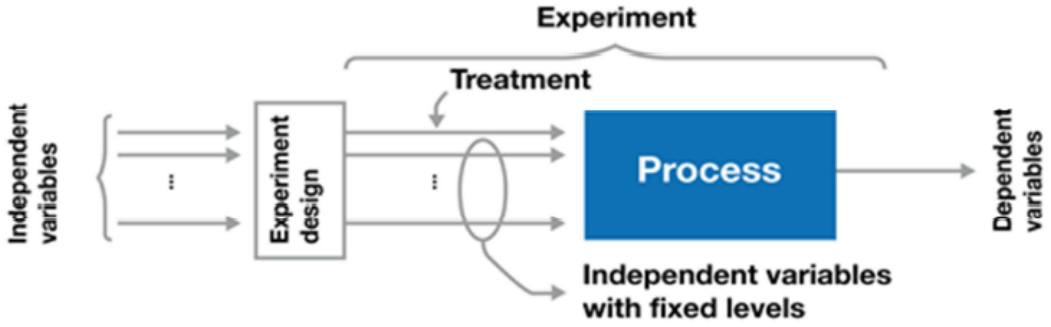

Figura 1.3 – Processo sperimentale

- *Variabile dipendente*: l'output del processo sperimentale. Quella variabile che si desidera studiare per investigare gli effetti dell'applicazione delle variabili indipendenti;
- *Variabili indipendenti*: tutte le variabili che sono manipolate e controllate al fine di produrre un effetto sulla variabile dipendente (risultato osservato);
- *Fattore*: caratteristiche per cui deve essere studiata la loro influenza sui risultati dell'esperimento (è un tipo specifico di variabile indipendente);
- *Trattamento*: uno dei possibili valori assunti dal fattore;
- *Unità sperimentale*: è l'oggetto dell'obiettivo;
- *Soggetto sperimentale*: le persone che applicano ciò che si vuole sperimentare;
- *Test/trial/prova*: la combinazione di trattamento, soggetto e unità sperimentale. Ad esempio, un experimental trial: N persone (soggetti sperimentali) usano il metodo di sviluppo innovativo (un livello del trattamento) per sviluppare il programma A (unità sperimentale).

4. LINEE GUIDA PER L'ESECUZIONE DI UNA RICERCA EMPIRICA

Se da una parte, la definizione dell'indagine ne determina le fondamenta: il perché eseguire l'indagine; la pianificazione specifica come essa deve essere eseguita. Una corretta pianificazione assicura un controllo sull'indagine durante tutta la sua esecuzione. In particolare, il processo prevede fasi che vanno dalla definizione, alla pianificazione ed esecuzione, alla analisi ed interpretazione dei dati raccolti, e infine al confezionamento dei risultati in pacchetti di conoscenza da essere riusati in futuro, per la replica della indagine o per la pianificazione di altre indagini.

Le considerazioni precedentemente esposte sono riassunte nella Figura 4.1.

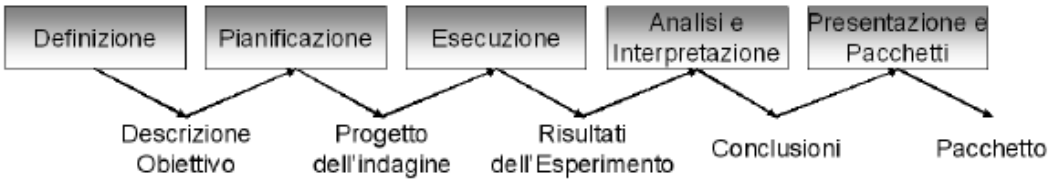

Figura 4.1 – Processo per la esecuzione di una ricerca empirica

In ciascuna fase possono essere assunti alcuni accorgimenti per eseguire accuratamente la ricerca. Di seguito si descrivono in ogni fase gli accorgimenti che provengono da una raccolta di esperienze.

4.1 Esperimento controllato

L'esperimento controllato è, in genere, chiamato semplicemente esperimento. Esso è eseguito in vitro. Perciò, questo tipo di esperimento consente un alto livello di controllo. L'obiettivo è quello di manipolare una o più variabili indipendenti e valutare tutte le variabili dipendenti che consentono di rispondere ai quesiti che costituiscono l'ipotesi di ricerca. Gli effetti di tale manipolazione vengono misurati, e in seguito tali misurazioni sono la base di analisi statistiche.

Un esperimento controllato è un'investigazione formale, rigorosa e controllata. I fattori chiave in esperimenti di questo genere possono essere identificati e manipolati. In tali esperimenti si ha il pieno controllo delle variabili coinvolte, quindi esse possono assumere i valori più adatti allo scopo della sperimentazione. L'attenzione è rivolta a distinguere e a confrontare i risultati di due contesti: un contesto di controllo (cioè quei

gruppi che lavorano con le tecnologie consolidate nell'impresa) e un contesto di investigazione (quei gruppi che lavorano con le tecnologie da introdurre). Gli esperimenti controllati sono utilizzati quando si vuole il pieno controllo dell'esperimento e ne si vuole manipolare il comportamento in maniera diretta, precisa e sistematica. Inoltre, questi esperimenti prevedono diversi modi di valutare i risultati. Infatti, è possibile effettuare sperimentazioni sia in contesti controllati, per esempio in un laboratorio dove gli eventi sono organizzati al fine di simulare una situazione reale, sia in contesti reali, cioè è eseguita una sperimentazione in campo sotto condizioni reali. Il livello di controllo in contesti reali è più difficile poiché sarà possibile avere sotto controllo solo alcuni dei fattori di interesse.

Un'importante assunzione della Ricerca Empirica nella Ingegneria del Software e di determinare quale tecnologia di sviluppo software dispiegare, cosa gli sviluppatori devono usare ed in quale situazione; in altre parole, l'interesse consiste nel paragonare varie tecnologie e competenze e nel determinare i loro effetti sullo sviluppo software. L'esperimento è un tipo di ricerca che può essere applicato, più o meno, direttamente per effettuare tali paragoni e misurare i relativi effetti. Gli esprimenti sono idonei per la descrizione causale, cioè "la descrizione delle conseguenze attribuibili ad un trattamento deliberatamente modificato".

4.2 Definizione

La prima attività prevede la descrizione dell'obiettivo dello studio. Si conviene che di seguito si usi il template Goal Question Metrics (GQM) [van Solingen] per tale descrizione. In breve, il template è composto di cinque parti:

- Analizzare <oggetto di studio> (dà il target dell'esperimento);
- Allo <scopo dello studio> (indica il motivo dello studio);
- Rispetto a <fattore di qualità> (di riferimento per definire l'ipotesi);
- Dal punto di vista <prospettiva> (dettaglia il punto di vista rispetto a cui si valuta l'obiettivo);
- Nel contesto <contesto> (di riferimento per la validità dei risultati).

Nel documento **Definizione del Modello di Qualità e Piano di Misurazione CS1** è descritto più dettagliatamente questo template.

La **Error! Reference source not found.** sottostante fornisce alcuni esempi di obiettivi di una indagine empirica.

Oggetto di studio	Scopo	Fattore di qualità	Prospettiva	Contesto
Prodotto	Caratterizzare	Efficacia	Sviluppatore	Soggetti
Processo	Monitorare	Costo	Manutentore	Oggetti
Modello	Valutare	Affidabilità	Project Manager	
Metrica	Prevedere	Manutenibilità	Manager d'impresa	
Teoria	Controllare	Portabilità	Committente	
	Cambiare		Utente/Ricercatore	

Tabella 4.1: Template GQM per la definizione dell'obiettivo

4.3 Contesto sperimentale

Il contesto sperimentale potrebbe descrivere le circostanze in cui una ricerca empirica è stata concepita, progettata ed eseguita oppure le circostanze in cui una nuova tecnologia software è stata sviluppata.

La descrizione del contesto assicura che:

- gli obiettivi della ricerca siano definiti appropriatamente
- la descrizione della ricerca sia sufficientemente dettagliata perché possa essere compresa, in tutte le sue componenti, da ricercatori e professionisti diversi da quelli che l'hanno eseguita. La descrizione del contesto riveste particolare importanza nello studio di osservazione e nell'esperimento formale.

Studio di osservazione

Gli studi empirici di osservazione o misurazione hanno una grande quantità di informazioni contestuali che devono essere comprese e che possono influenzare i risultati. I ricercatori devono identificare i fattori che possono concernere la generalità e l'utilità delle conclusioni. Esempi di tali fattori possono essere:

- il settore produttivo in cui erogare servizi (banche, telecomunicazioni, turismo, etc.);
- le competenze e le esperienze del team coinvolto nella gestione ed erogazione del servizio (nel linguaggio, nel metodo, nel dominio applicativo);
- il tipo di richieste da evadere;
- processi di gestione del servizio (processi di risoluzione, controllo, etc.).

Non esistono modalità accreditate o standard per individuare quali informazioni contestuali è necessario rilevare. In ogni caso queste servono a paragonare o contrastare studi simili. Per raggiungere tale scopo è necessario approfondire la ricerca in almeno due aree:

- identificazione di fattori fuorvianti per particolari tipi di ricerca;
- specificazione di una tassonomia o una ontologia del contesto.

Esperimento formale

È necessario che nella sperimentazione di nuove tecniche (ispezioni, pattern di progetto) queste non siano eccessivamente semplificate. L'accademia tende a semplificare le tecniche da sperimentare, mentre l'industria preferisce dettagliare ogni tecnica. In particolare, si deve essere sicuri che si sta sperimentando l'ultima versione della tecnica con tutti gli accorgimenti migliorativi che l'esperienza d'uso della stessa ha evidenziato ed indotto a realizzare. Inoltre, devono essere misurati tutti gli aspetti che possono apportare benefici o problemi all'adozione della tecnica.

4.4 Pianificazione

La pianificazione consente di individuare le misure necessarie per la ricerca durante tutta la sua esecuzione. Questa fase comprende la selezione del

contesto in cui eseguire la ricerca, la formulazione delle ipotesi sperimentali, la selezione dei soggetti, e quindi degli individui, che dovranno essere coinvolti nella ricerca e l'identificazione delle variabili dipendenti ed indipendenti. Infine, comprende la specifica degli strumenti da adottare per l'esecuzione vera e propria. Una possibile sintesi grafica di questa fase può essere rappresentata come segue:

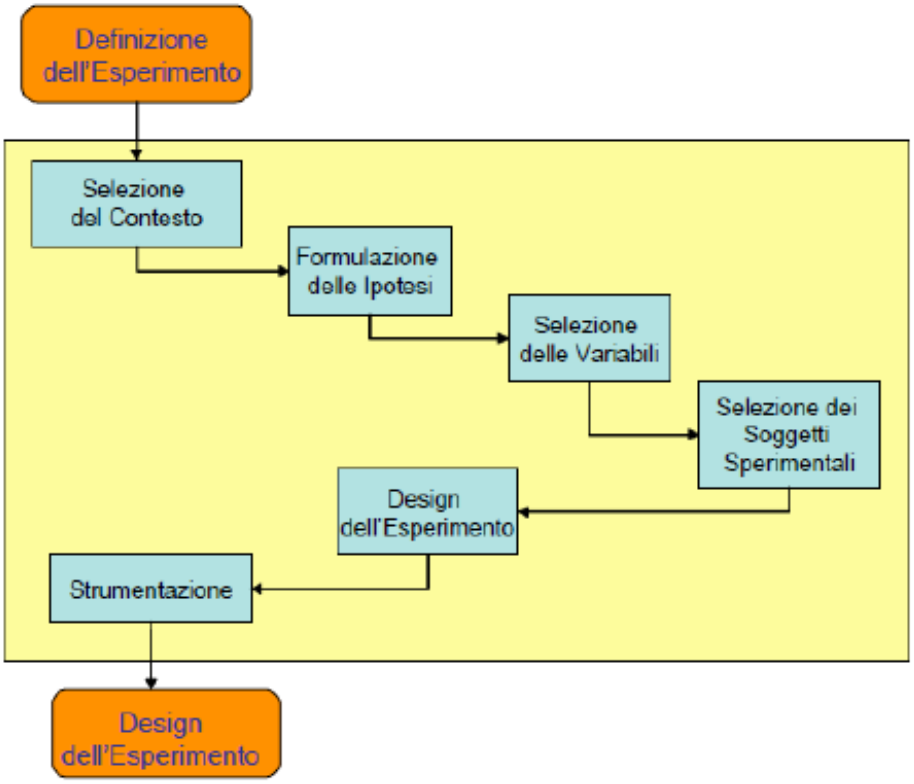

Figura 4.2 – Attività della fase di pianificazione

4.5 Ipotesi sperimentali

La base per le analisi statistiche di un esperimento è il test delle ipotesi. Nella fase di pianificazione, la definizione dell'esperimento è formalizzata in ipotesi.

L'ipotesi è formalizzata in:

- *ipotesi nulla*, tale ipotesi afferma che non ci sono trend o pattern nei risultati dell'esperimento; le differenze rilevate nelle osservazioni sono casuali. Questa è l'ipotesi che lo sperimentatore desidera

rigettare con un più alto livello di significatività possibile. Un esempio di ipotesi nulla è che una nuova metodologia sia equivalentemente efficace a quella attualmente in uso.

- *ipotesi alternativa*, è l'ipotesi rispetto alla quale la ipotesi nulla è rigettata. Un esempio di ipotesi alternativa è che una nuova metodologia sia più efficace di quella attualmente in uso.

La letteratura fornisce diversi test statistici per valutare i risultati di un esperimento. Tali test richiedono che le ipotesi siano formulate prima della selezione e della esecuzione dei test statistici.

Il test delle ipotesi implica vari tipi di rischi: il test può rigettare una ipotesi vera oppure il test può non rigettare una ipotesi falsa. Tali rischi sono indicati come errore-tipo-I ed errore-tipo-II. In particolare:

- ***Errore-tipo-I (di prima specie)***: consiste nel rifiutare l'ipotesi nulla quando è vera. Ovvero, la probabilità di commettere un errore di tipo I è espresso come: P(errore-tipo-I) = P (Rigettare H_0 | H_0 vera). Ad esempio, l'errore-tipo-I è la probabilità di rigettare H_0 anche se in media le due metodologie considerate sono equivalentemente efficaci.
- ***Errore-tipo-II (di seconda specie)***: consiste nel non rifiutare l'ipotesi nulla quando è falsa. In altre parole, la probabilità di commettere un errore di tipo II è espressa come: P (errore-tipo-II) = P (Non Rigettare H_0 | H_0 falsa). Ad esempio, l'errore-tipo-II è la probabilità di non rigettare l'ipotesi nulla H_0 anche se in media le due metodologie hanno medie differenti.

Test di ipotesi		Decisione	
		Non rifiuto H_0	Rifiuto H_0
Stato reale	H_0 è vera	Decisione corretta	Errore di prima specie

	H_0 è falsa	Errore di seconda specie	Decisione corretta

Tabella 4.2 – Decisioni e tipi di errore

Quando una ipotesi non ha un'esplicita teoria di derivazione, risulta essere solo un test. Per esempio, se si volesse conoscere l'esistenza di una relazione tra numero ciclomatico e numero di fault in un modulo, l'ipotesi *"non c'è correlazione tra numero ciclomatico e numero di fault"* è superficiale perché non riflette alcuna teoria esplicativa. Ovvero la risposta a questa ipotesi non aumenta la conoscenza nell'ingegneria del software.

Infatti, senza una sottostante teoria non è possibile comprendere la ragione per cui i risultati di una ricerca empirica sono inconsistenti. Quando si hanno risultati consistenti, come nel caso di molti studi sulla ispezione, non si può asserire come migliorare oppure ottimizzare il risultato se non comprendiamo perché il fenomeno si verifica. Per esempio, si potrebbe usare la teoria cognitiva secondo la quale un sistema con più che 5 punti di decisioni binarie risulta essere difficilmente governabile e comprensibile da un umano. Da questa teoria possono discendere le seguenti ipotesi *"il numero ciclomatico non ha impatto sulla correttezza di un modulo"*; *"il numero ciclomatico non ha correlazione con la comprensibilità del codice"*; *"il numero ciclomatico non ha correlazione con la manutenibilità del codice"*. Quando queste ipotesi sono negate ed il numero ciclomatico è minore di 5 è confermata la teoria precedente. Se le ipotesi sono confermate ed il numero ciclomatico è minore di 5 allora si potrebbe provare a modificare la teoria, diminuendo il numero soglia per la comprensibilità di un codice. Nel caso in cui, nonostante le modifiche, i risultati della ricerca sono inconsistenti, allora la validità della teoria sarebbe messa in dubbio almeno per l'ingegneria del software.

4.6 Selezioni delle variabili

Prima che un qualsiasi esperimento possa essere progettato, è indispensabile identificare le variabili dipendenti ed indipendenti.

Le variabili indipendenti sono quelle variabili che si possono controllare e modificare nel corso dell'esperimento. La scelta di tali variabili richiede la conoscenza del dominio applicativo in cui si sta svolgendo la ricerca. Tali variabili dovrebbero avere un qualche effetto sulle variabili dipendenti ed è necessario che siano controllabili. La scelta di tali variabili richiede che siano

specificate anche la scala di misurazione, i possibili valori assumibili, e quando tali valori dovranno essere rilevati.

L'effetto del trattamento è misurato sulle variabili dipendenti. In genere tale variabile non è osservabile direttamente, ma è una metrica calcolata attraverso un modello di calcolo che implica la raccolta di metriche osservate durante la esecuzione della ricerca. La definizione delle variabili dipendenti implica anche la specifica della scala di misurazione e dei valori rilevabili e attesi.

4.7 Selezioni dei soggetti sperimentali

La selezione dei soggetti sperimentali è un aspetto rilevante nella progettazione di una ricerca empirica. La selezione è strettamente collegata alla generalizzazione dei risultati dell'esperimento. Per generalizzare i risultati nell'ambito di una popolazione, il campione deve essere rappresentativo della popolazione stessa.

La popolazione deve essere esattamente definita altrimenti è impossibile effettuare alcuna inferenza dai risultati degli esperimenti. Infatti, se non è ben definita la popolazione dalla quale sono stati scientificamente estratti i soggetti o gli oggetti sperimentali è impossibile conoscere entro quale popolazione sono generalizzabili i risultati della ricerca.

Uno dei modi più convincenti di selezione è l'estrazione casuale dei soggetti o degli oggetti dall'intera popolazione. Spesso è opportuno definire i criteri di inclusione ed eventualmente anche quelli di esclusione.

Se si usa l'estrazione casuale, è opportuno dichiarare gli strumenti utilizzati per l'estrazione. Se, invece, si usa un altro metodo si deve descrivere il metodo ed il razionale che deve convincere circa la sua capacità di selezionare un campione significativo.

Per esempio, per valutare l'efficacia di una tecnologia per la progettazione software si possono selezionare professionisti che hanno una buona esperienza nella progettazione. Si escludono tutti coloro che hanno meno di 1800 ore di impegno nella progettazione perché considerati poco esperti per evidenziare l'efficacia della tecnologia e tutti coloro che hanno più di 5400 ore perché troppo esperti e potrebbero far risultare efficace la tecnologia non per le caratteristiche della stessa ma per la loro esperienza che gli potrebbe consentire di produrre buoni progetti qualunque sia la tecnologia che usano.

4.8 Design sperimentale

Il progetto dell'indagine empirica descrive le azioni che devono essere eseguite e gli strumenti da utilizzare. Il progetto descrive anche i vincoli temporali con cui le stesse attività devono essere eseguite.

Nella progettazione si tenta di prevedere tutti i problemi che si possono verificare durante la ricerca. In particolare, tutte le così dette minacce. Per ognuna di esse è necessario prevedere una soluzione che la elimini o che consenta di isolarla in modo che si possano individuare nettamente le relazioni causa-effetto senza che la minaccia si sostituisca al reale effetto od alla reale causa.

Potrebbe capitare che per il superamento di qualche minaccia esistano soluzioni possibili in concorrenza. In questo caso il progettista deve scegliere una delle possibili soluzioni seguendo un preciso razionale.

I soggetti e gli oggetti devono essere rappresentativi della popolazione altrimenti non è possibile generalizzare i risultati della ricerca.

Uno dei modi più convincenti di selezione è l'estrazione casuale dei soggetti o degli oggetti dall'intera popolazione. Spesso è opportuno definire i criteri di inclusione e, se è il caso, anche quelli di esclusione.

Se si usa l'estrazione casuale, è opportuno dichiarare gli strumenti utilizzati per l'estrazione. Se, invece, si usa un altro metodo si deve descrivere il metodo ed il razionale che deve convincere circa la sua capacità di selezionare un campione significativo. Per esempio, per valutare l'efficacia di una tecnologia per la progettazione software si possono selezionare professionisti che hanno una buona esperienza in questo ambito. Si escludono tutti coloro che hanno meno di 1800 ore di impegno nella progettazione perché considerati poco esperti per evidenziare l'efficacia della tecnologia e tutti coloro che hanno più di 5400 ore perché troppo esperti e potrebbero far risultare efficace la tecnologia non per le caratteristiche della stessa ma per la loro esperienza che gli potrebbe consentire di produrre buoni progetti qualunque sia la tecnologia che usano. È opportuno utilizzare schemi sperimentali semplici oppure che siano approfonditi dalla letteratura statistica. Nel caso che si ritiene necessario utilizzare schemi progettuali ed analisi che non siano accreditati dalla letteratura è opportuno consultare uno statistico esperto.

Ogni schema sperimentale ha le sue implicazioni in termini di minacce e di problemi. Se uno schema è semplice, risulta possibile governarlo con

l'esperienza dello sperimentatore, anche quando non ci sono esperienze pregresse. Se uno schema non è semplice, risulta opportuno utilizzare esperienze pregresse e, quindi, far riferimento alla letteratura statistica. Dalla capacità di governare uno schema sperimentale dipende la probabilità che un ricercatore generi errori.

Per completezza, alcuni possibili errori che il ricercatore può fare nella sperimentazione di uno schema sperimentale sono:

- *Produzione di schemi annidati* dove l'unità sperimentale è un gruppo di soggetti piuttosto che un soggetto individuale e non considerare il radicale cambio della natura della sperimentazione e delle modalità per analizzarlo. Per esempio: supposto che uno sperimentatore abbia 4 gruppi di 10 studenti e vuole valutare la differenza di efficacia di due tecnologie, se assegna ogni studente ad una tecnologia, casualmente, egli ha 38 gradi di libertà per verificare la differenza di efficacia; se egli assegna un gruppo, come un insieme ad ogni tecnologia allora ha solo 2 gradi di libertà. In generale, a maggiore numero di gradi di libertà corrisponde maggiore qualità del progetto.
- *Uno schema complesso* può indurre in errori di analisi dei dati. Un caso è l'uso di schemi ad incrocio tra soggetti sperimentali e trattamenti, per ridurre il problema di un piccolo numero di soggetti sperimentali. In questo schema, uno stesso soggetto sperimentale è soggetto a differenti trattamenti; allora è necessario analizzare ed isolare il rischio di maturità che un soggetto porta in un trattamento avendola acquisita nel trattamento precedente. Ad esempio, si supponga di confrontare due tecniche di ispezione: uno stesso soggetto sperimentale quando sperimenta la seconda tecnica di ispezione adopera esperienza e modalità che ha acquisito nell'uso della prima

tecnica. Normalmente questo problema può essere fronteggiato con la scelta casuale o con il bloccaggio di alcune variabili.

È importante giustificare la scelta delle misure dell'esito dell'esperimento in termini della loro rilevanza per gli obiettivi dello studio empirico. Più precisamente, la rilevanza delle misure dell'esito rispetto agli obiettivi della ricerca è essenziale per dimostrare che il progetto dell'indagine sia appropriata agli stessi obiettivi della ricerca. Questo è particolarmente richiesto quando si usano misure indirette dell'esito. Per esempio, se si dovesse valutare l'impegno uomo nella manutenzione, utilizzando una nuova tecnologia e si vuole misurare questo con il numero delle LOC modificate, è necessario esplicitare come e perché questa misura valuta correttamente l'esito dell'esperimento. Un aspetto fondamentale circa i risultati di un esperimento, da tenere conto durante la identificazione del design sperimentale, concerne la loro validità. I risultati sperimentali hanno una adeguata validità se sono validi per la popolazione rispetto alla quale vanno generalizzati. La letteratura presenta diversi schemi per classificare le minacce che possono intaccare la validità dei risultati.

In generale si possono individuare cinque tipologie di validità sperimentali, per ciascuna delle quali esistono possibili minacce da tenere in considerazione durante l'indagine:

1) Validità delle Conclusioni Statistiche: la relazione tra il trattamento ed il risultato deve avere significatività statistica. In questo caso le minacce individuate sono:
    - potenza statistica, debolezza o inadeguatezza dei test statistici;
    - assunzioni: correttezza delle assunzioni fatte per i test statistici;
    - fishing: individuazione nei risultati di effetti che non hanno adeguato livello di significatività;
    - affidabilità delle misure;
    - affidabilità della realizzazione del trattamento;
    - rumori nell'assetto dell'esperimento;

- eterogeneità dei soggetti sperimentali che non sono colmabili con la casualità della distribuzione degli stessi soggetti.

2) Validità Interna: la relazione tra il trattamento ed i risultati deve essere causale cioè quando si può provare che le modifiche apportate alla variabile indipendente causino quelle rilevate nella variabile dipendente. Possibili minacce sono rappresentate da:
   - *Storia*: trattamenti differenti effettuati sullo stesso soggetto in tempi diversi possono avere risultati diversi solo perché sono diverse le condizioni in cui opera il soggetto sperimentale. Per esempio, un trattamento fatto in un giorno prima di andare in vacanza ed un altro nel giorno successivo alla vacanza, potrebbe aver risultati differenti solo perché le condizioni del soggetto sono diverse in termini di concentrazione sulle attività. E' opportuno svolgere l'esperimento in uno stesso giorno od in differenti giorni con le stesse caratteristiche ed utilizzare gruppi di verifica.
   - *Maturità*: uno stesso trattamento ha effetti cangianti durante un esperimento solo perché la sua esecuzione richiede un certo tempo. Esempi negativi sono rappresentati da noia o stanchezza mentre un esempio positivo è l'apprendimento. E' opportuno, bilanciare il tempo di durata di un esperimento ed utilizzare gruppi di verifica.
   - *Interazione tra prove successive*: nel caso in cui vengano eseguite più prove su uno stesso soggetto sperimentale, la prova successiva può dare risultati diversi perché il soggetto ha

appreso gli errori che ha fatto nella prova precedente. Risulta opportuno non far conoscere i risultati delle prove precedenti.

- *Strumentazione*: il materiale disponibile per l'esperimento ha dei faults, oppure gli strumenti hanno delle failures che possono influenzare negativamente gli esperimenti. E' opportuno preparare adeguatamente il materiale per l'esperimento e verificarlo prima di eseguire l'esperimento.
- *Regressione statistica*: i soggetti scelti sulla base di una precedente valutazione potrebbero influenzare il risultato attraverso le caratteristiche dei soggetti sperimentali. Ad esempio: i migliori 10 costituiscono un gruppo ed i peggiori 10 costituiscono un altro gruppo. È opportuno non basare la selezione su una o più valutazioni a meno che il progetto dell'esperimento non preveda che i soggetti presentino uno specifico parametro di competenza. In ogni caso, è opportuno utilizzare almeno un gruppo di verifica.
- *Selezione*: la tecnica di selezione dei soggetti sperimentali potrebbe scegliere i soggetti in modo che non siano rappresentativi dell'intera popolazione. Per esempio, selezionare i volontari potrebbe significare di avere tra i soggetti sperimentali solo le persone motivate. È opportuno utilizzare gruppi di verifica.
- *Mortalità*: la fuoriuscita di alcuni soggetti sperimentali potrebbe mettere in discussione la rappresentatività della popolazione con il gruppo di soggetti sperimentali. È opportuno utilizzare gruppi di verifica.

3) Validità di Costrutto. Il trattamento riflette bene le cause ed il risultato riflette bene gli effetti. Le minacce si identificano in:
    - Inadeguata definizione del costrutto che potrebbe aver indotto ad eseguire misure poco significative per lo scopo dell'esperimento. Spesso è indotto da termini generici nella definizione del quesito di ricerca.
    - Oggetto sperimentale unico che non esprime il costrutto; per esempio, un solo documento o un solo metodo.
    - Misura o osservazione di un unico tipo che possa orientare il risultato.
    - Confusione tra costrutti e livello dei costrutti: qualche caratteristica che dovrebbe essere definita come parametro nel costrutto è confusa con il livello della stessa caratteristica. Ad esempio, la presenza o l'assenza di esperienza in un linguaggio è confusa con la valutazione del livello di esperienza dei soggetti sperimentali.
    - Interazione tra trattamenti differenti: effetto di un trattamento effettuato sui soggetti sperimentali che influenza il comportamento di questi ultimi quando sono sottoposti al successivo trattamento.
4) Validità Sociale: Lo sperimentatore ed i soggetti sperimentali si comportano in modo da influenzare i risultati. Le minacce sociali alla validità comprendono:
    - Influenza sulle ipotesi: i soggetti sperimentali immaginano quali sono le ipotesi che stanno alla base dell'esperimento e si comportano in favore o contro la conferma delle ipotesi.

- Apprensione per la valutazione: i soggetti sperimentali, preoccupati per la valutazione che ne deriverà dai loro risultati nell'esecuzione dell'esperimento, falsano il loro comportamento per avere una migliore valutazione.
- Aspettative dello sperimentatore: la esecuzione dell'esperimento può essere influenzata, consciamente od inconsciamente dalle aspettative dello sperimentatore. Questa minaccia si può evitare facendo eseguire l'esperimento da un ricercatore diverso a quello interessato al risultato dell'esperimento.

5) Validità Esterna. I risultati possono essere generalizzati. Possibili minacce da tenere in conto includono:
   - Rappresentatività dei soggetti: i soggetti sperimentali potrebbero essere insufficienti per la caratterizzazione della popolazione degli sviluppatori
   - Rappresentatività dell'ambiente sperimentale: l'ambiente sperimentale non è rappresentativo del contesto in cui saranno trasferiti i risultati dell'esperimento. Ad esempio nell'esperimento si usano strumenti obsoleti rispetto a quelli che, normalmente, si usano nell'impresa.
   - Stagionalizzazione dell'esperimento: l'esecuzione dell'esperimento coincide con un evento che potrebbe alterare il risultato.

4.9 Progetto dell'indagine

Il progetto della ricerca descrive le azioni che devono essere eseguite e gli strumenti da utilizzare. Il progetto descrive anche i vincoli temporali con cui le stesse attività devono essere eseguite.

Nella progettazione si tenta di prevedere tutti i problemi che si possono verificare durante la ricerca. In particolare, tutte le così dette *minacce*. Per ognuna di esse è necessario prevedere una soluzione che la elimini o che consenta di isolarla in modo che si possano individuare nettamente le relazioni causa-effetto senza che la minaccia si sostituisca al reale effetto od alla reale causa.

Potrebbe capitare che per il superamento di qualche minaccia esistano soluzioni possibili in concorrenza. In questo caso il progettista deve scegliere una delle possibili soluzioni seguendo un preciso razionale.

1. *Definire il processo attraverso il quale si assegnano i soggetti o gli oggetti ai trattamenti sperimentati.*

   I soggetti e gli oggetti devono essere assegnati ai trattamenti in modo che non possano influenzare il risultato. Spesso questa associazione è effettuata casualmente. Se i soggetti sono pochi, anche questa tecnica non garantisce la mancanza di sbilanciamento; per esempio, se i soggetti sono pochi è possibile che i due programmatori più esperti della popolazione disponibile siano assegnati ad uno stesso trattamento e questo influenzerebbe il risultato.

   Qualche volta è necessario utilizzare tecniche di randomizzazione più complesse. Ad esempio, pre-classificare i soggetti per la loro preparazione oppure per il sesso di appartenenza ed assicurarsi che ad ogni trattamento siano assegnati un numero bilanciato estratto da ognuna delle classi prodotte.

2. *E' opportuno utilizzare schemi sperimentali semplici oppure che siano approfonditi dalla letteratura statistica.*

   Ogni schema sperimentale ha le sue implicazioni in termini di minacce e di problemi. In presenza di uno schema semplice è possibile governarlo con l'esperienza dello sperimentatore, anche quando non ci sono esperienze pregresse. Se uno schema non è semplice, è opportuno utilizzare esperienze pregresse e quindi far riferimento alla

letteratura statistica. Dalla capacità di governare uno schema sperimentale dipende la probabilità che un ricercatore generi errori. Per completezza, alcuni possibili errori che il ricercatore può fare nella sperimentazione di uno schema sperimentale sono:

- Produzione di schemi annidati dove l'unità sperimentale è un gruppo di soggetti piuttosto che un soggetto individuale e non considerare il radicale cambio della natura della sperimentazione e delle modalità per analizzarlo.
- Uno schema complesso può indurre in errori di analisi dei dati. Un caso è l'uso di schemi ad incrocio tra soggetti sperimentali e trattamenti, per ridurre il problema di un piccolo numero di soggetti sperimentali. In questo schema, uno stesso soggetto sperimentale è soggetto a differenti trattamenti; allora è necessario analizzare ed isolare il rischio di maturità che un soggetto porta in un trattamento avendola acquisita nel trattamento precedente. Normalmente questo problema può essere fronteggiato con la scelta casuale o con il bloccaggio di alcune variabili. In ogni caso l'analisi dei risultati risulta essere più complessa .

3. *Definire correttamente l'unità sperimentale*.

Onde evitare di commettere errori sulla definizione dell'unità sperimentale è necessario definirla precisamente quando si esegue una procedura di randomizzazione. Questo errore è spesso commesso nelle indagini di ricognizione. Quando si invia un questionario ad un insieme di imprese, l'unità sperimentale è l'impresa ma il soggetto sperimentale è individuale, colui che risponde al questionario. Se questa distinzione non è chiara, il processo di randomizzazione sarà inflazionato e quindi l'analisi dei dati è a sua volta incorrettamente inflazionata.

4. *Gli esperimenti formali hanno bisogno di pre-esperimenti o il pre-calcolo ad identificare o stimare la dimensione minima richiesta del campione sperimentale.*

   La dimensione del campione sperimentale consente di determinare la probabilità di trovare correttamente la differenza (rigettando l'ipotesi nulla) quando essa esiste. Per un buon esperimento di conferma i ricercatori si aspettano di avere informazioni circa la misura della differenza tra gli effetti del trattamento e la varianza delle unità sperimentali. Da questa differenza si può calcolare il numero di soggetti sperimentali, o più correttamente di unità sperimentali, richiesti per avere una data probabilità di rigettare l'ipotesi nulla quando essa è falsa. Questa probabilità è chiamata *potenza del test*. Normalmente la progettazione dell'esperimento mira ad avere una potenza pari a 0,8.

5. *Usare un appropriato livello di schermatura.*

   Nella ricerca empirica è necessario schermare l'influenza delle attese del soggetto sperimentale o del ricercatore sui risultati. Negli esperimenti dell'ingegneria del software il problema è rilevante perché i soggetti sperimentali possono fortemente influenzare il risultato con il loro comportamento. Per esempio, chi è contrario ad una tecnologia ed è a conoscenza che dai risultati della sperimentazione dipenderà la decisione circa l'adozione della stessa, si comporterà (durante la sperimentazione) in modo tale che i risultati della stessa risultino sfavorevoli alla tecnologia. Ci sono, però, molti livelli di schermatura possibile:

   - Schermatura dell'allocazione dei materiali. Questo implica che oltre all'assegnazione dei soggetti al trattamento deve essere randomizzata anche l'allocazione dei materiali che i soggetti sperimentali devono utilizzare. Per esempio, si consideri di provare due tecniche di test in cui gli oggetti sperimentali sono rappresentati da un insieme di programmi e ognuno di essi sia testato con entrambe le tecniche: prima si assegna casualmente il

soggetto sperimentale alla tecnica di teste, dopo si assegna, altrettanto casualmente, il programma che ogni soggetto sperimentale deve verificare con la tecnica di test che gli è stata assegnata.

- Schermatura della tracciabilità. Quando uno dei risultati di ricerca deriva da risposte a quesiti sottoposti ai soggetti sperimentali, è opportuno che il formato delle risposte non sia dipendente dal trattamento. In tal modo l'elaborazione delle risposte non sarebbe influenzata dall'aspettativa che l'analizzatore ha su uno specifico trattamento. Ad esempio, se si volessero confrontare due tecniche di test ed il soggetto sperimentale deve descrivere i difetti trovati, il modulo su cui si descrivono i difetti deve essere lo stesso per entrambi i tipi di test. In questo modo solo dopo l'analisi dei difetti rilevati si potrà risalire al soggetto che lo ha prodotto attraverso il codice su ogni modulo, e quindi, alla tecnica di test utilizzata.
- Schermatura dell'analisi. L'analisi dei dati deve essere effettuata senza sapere da quale gruppo di trattamento i dati provengano. Questo si può ottenere codificando l'insieme dei dati con un codice che individua il trattamento e che può essere decodificato solo dopo che l'analisi è stata eseguita. Questa schermatura è ritenuta da molti statistici un buon mitigatore del rischio di "fishing" per i risultati.

6. *Se un ricercatore non può evitare di valutare il suo stesso lavoro allora deve esplicitare i suoi interessi nella sperimentazione e quali accorgimenti ha adottato per minimizzare l'influenza dei suoi interessi sul risultato.*

Uno sperimentatore potrebbe avere necessità di valutare una tecnologia di sua produzione con un'altra tecnologia o di valutare una tecnologia di cui si ritiene un sostenitore. In questo caso, egli definisce le ipotesi da verificare ed anche una preliminare valutazione dei risultati; egli non è, però, il migliore valutatore dei risultati perché il suo entusiasmo potrebbe influenzare l'analisi e la interpretazione dei risultati. Uno degli accorgimenti migliori in questo caso è l'affidamento della valutazione ad un soggetto indipendente.

7. *Evitare l'uso di controlli tranne che non sia possibile definirli senza alcuna ambiguità.*

   Questo problema è di grande rilevanza perché gran parte dei risultati di una ricerca sono fondati sul comportamento dell'uomo. In particolare, non è possibile valutare l'effetto di una tecnologia paragonando il suo uso al non uso della stessa perché è difficile controllare che un soggetto sperimentale non usi alcuna tecnologia, tuttalpiù utilizzerà la sua esperienza. È possibile invece solo confrontare l'efficacia di due tecnologie diverse, entrambe contenute nella stessa indagine. Questo fenomeno nella sperimentazione industriale implica che è possibile confrontare una nuova tecnologia rispetto a quella utilizzata normalmente dalla stessa impresa. Di conseguenza è difficile paragonare i risultati delle sperimentazioni per una stessa nuova tecnologia eseguite in due imprese differenti.

8. *Definire dettagliatamente i trattamenti da provare nella ricerca.*

   È necessario definire esattamente il protocollo del trattamento per poter replicare conformemente l'esperimento sia in laboratori sia in imprese differenti. La replica conforme consente di poter paragonare i risultati ricavati nei diversi esperimenti.

9. *Giustificare la scelta delle misure dell'esito dell'esperimento in termini della loro rilevanza per gli obiettivi dello studio empirico.*

   La rilevanza delle misure dell'esito rispetto agli obiettivi della ricerca è essenziale per dimostrare che il progetto dell'indagine sia

appropriata agli stessi obiettivi della ricerca. Questo è particolarmente richiesto quando si usano misure indirette dell'esito.

4.10 Strumenti

Nella esecuzione della ricerca empirica i soggetti sperimentali useranno i prototipi dimostratori sviluppati durante il progetto di ricerca.

4.11 Esecuzione

Operativamente, ogni soggetto sperimentale utilizzerà la soluzione innovativa, ossia il dimostratore. Saranno rilevate le variabili dipendenti, in seguito ad ogni trattamento e ad intervalli periodici, stabiliti prima dell'inizio dell'indagine.

I dati raccolti saranno poi analizzati per valutare l'efficacia della soluzione.

La fase di esecuzione è sintetizzata nella figura 4.3, in tre attività.

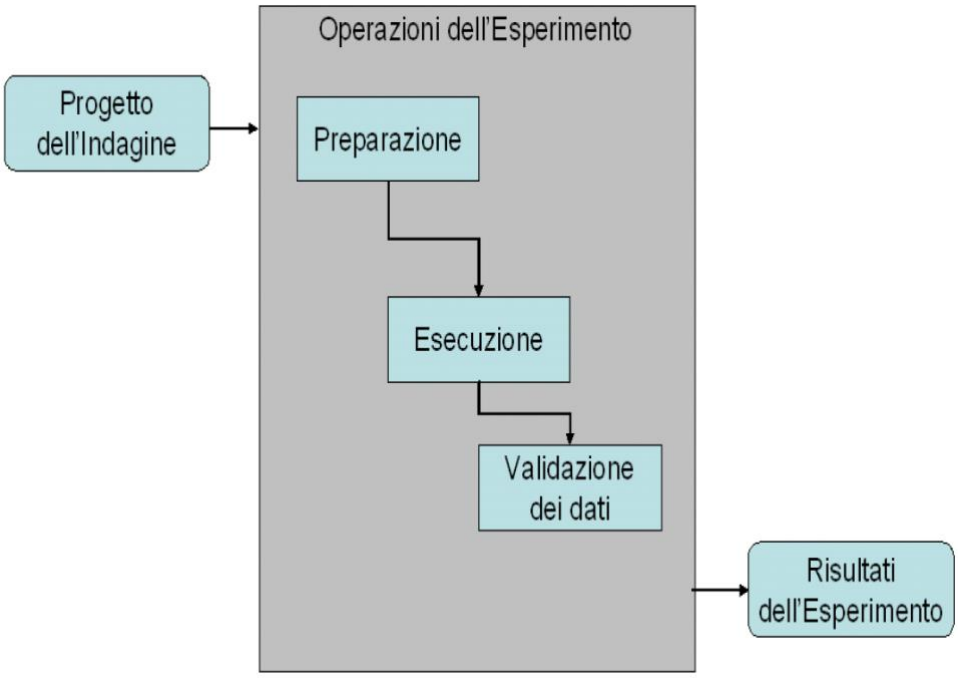

Figura 4.3 – Attività della fase di esecuzione

Durante la preparazione è necessario motivare i partecipanti ed ottenere il loro consenso, nonché predisporre tutta la strumentazione; segue poi

l'esecuzione del progetto dell'esperimento e la raccolta dei dati; questi saranno poi validati prima di poterli analizzare.

4.12 Analisi ed interpretazione

Dopo aver raccolto i dati durante la fase di esecuzione della ricerca, questi devono essere interpretati in modo da pervenire a conclusioni che consentano di provare e quindi accettare o meno le ipotesi sperimentali formulate precedentemente. In primo luogo, i dati raccolti sono sintetizzati attraverso statistiche descrittive per eventuali analisi qualitative esplicative degli stessi dati. È opportuno rappresentare graficamente i dati raccolti per individuare eventuali *outlier*, e decidere se è necessario escluderli o meno dal campione e quindi ridurre il data set. Infine, i dati sono analizzati attraverso il test delle ipotesi, in cui le ipotesi sperimentali sono validate statisticamente rispetto ad uno specifico livello di significatività. Le fasi dell'attività di analisi ed interpretazione sono illustrate in Figura 4.4.

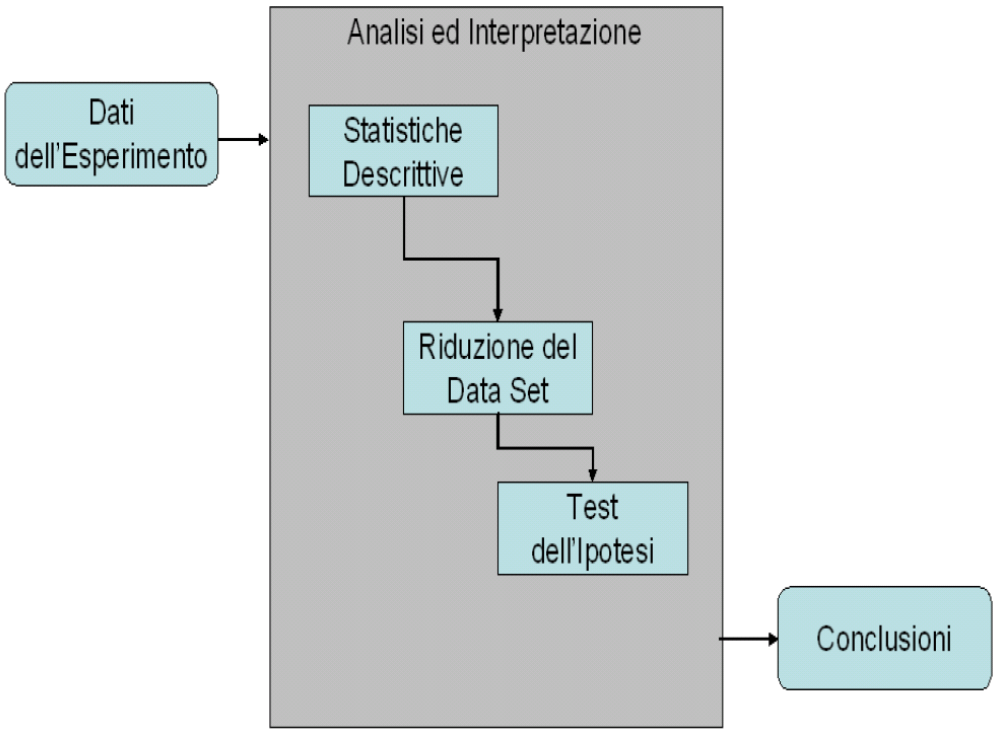

Figura 4.4 – Attività della fase di analisi dei risultati

Esistono essenzialmente due approcci per analizzare i risultati sperimentali:

- Analisi classica o frequentistica adottata dalla maggior parte dei package statistici.
- Analisi bayesiana che usa sistematicamente informazioni a priori. In genere tali informazioni possono essere ottenute da precedenti studi del fenomeno di interesse o dalle opinioni degli esperti.

Un altro aspetto di interesse generale è la scelta tra analisi parametriche e non parametriche. Se la distribuzione delle variabili può essere identificata, allora l'analisi parametrica è più potente di quella non parametrica. Invece è più appropriata l'analisi non parametrica quando la distribuzione delle variabili non è identificata; la maggior parte delle analisi non parametriche, in questo caso, sono più efficienti rispetto alle loro controparti parametriche; inoltre, sono più efficaci nel caso di campioni di piccole dimensioni.

Le linee guida per l'analisi statistica hanno lo scopo di assicurare che i risultati sperimentali siano analizzati correttamente. Fondamentalmente i dati sarebbero analizzati in accordo con il progetto dell'indagine. Perciò una ben definita progettazione dell'indagine consente di scegliere una corretta analisi.

Spesso si hanno pochi dati ed è difficile averne altri dello stesso tipo. In questo caso si tende a sovra usare il data set per effettuare un grande numero di test statistici. Quando si fanno molti test statistici su pochi dati si corre il rischio di ottenere risultati statisticamente significativi solo per caso. Un metodo per far fronte a questo rischio è l'aggiustamento della soglia di significatività del test.

Ad esempio, se sono disponibili 10 test indipendenti e si richiede un livello di significatività globale di 0,05, l'aggiustamento di Bonferroni richiede un livello di significatività pari a 0,005 per ogni test individuale. Rosemberger propone un aggiustamento meno severo ma in ogni caso ogni singolo test deve avere un livello di significatività molto più alto di quello che si desidera come significatività dell'insieme di test.

È necessario fare un'analisi di sensibilità prima di decidere le tecniche statistiche da utilizzare per la loro interpretazione. L'analisi di sensibilità serve a comprendere come i singoli dati o cluster si relazionano al comportamento dell'intera collezione. Per esempio, la presenza di outlier in un insieme di dati rende inefficace qualunque interpretazione della tendenza

centrale dello stesso insieme. I possibili tipi di analisi di sensibilità suggeriti sono:

- identificare gli outlier;
- identificare e trattenere i dati che sono influenti, che non sono necessariamente coincidenti con gli outlier;
- per la regressione lineare, valutare l'indipendenza delle variabili.

E' sempre necessario assicurarsi che i dati abbiano caratteristiche conformi ai test che si applicano su di essi. Spesso è possibile rendere conforme la serie di dati ai test attraverso opportune trasformazioni degli stessi. Molti package statistici forniscono queste trasformazioni di adeguamento.

E' necessario controllare la validità dei dati prima di analizzarli.

L'esame grafico dei dati spesso aiuta ad individuare difetti nei dati stessi. Un'altra tecnica è l'uso di analisi esplorative prima di effettuare analisi di maggiore approfondimento.

Molte analisi sono effettuate dai package statistici. È opportuno non affidarsi completamente a queste analisi. È consigliabile, che programmi diversi effettuino le stesse analisi per confrontare i risultati.

4.13 Statistiche descrittive

Le statistiche descrittive servono a presentare il campione di dati sotto osservazione sintetizzandone gli aspetti di maggiore interesse:

- tendenza centrale o concentrazione / localizzazione dei dati;
- dispersione / concentrazione dei dati;
- dipendenza.

Sono solitamente utilizzate prima dell'esecuzione di test delle ipotesi al fine di meglio comprendere la natura dei dati osservati ed identificare eventuali dati anomali (outlier). La rappresentazione grafica è una modalità utile per avere una overview sul data set. Alcuni esempi di grafici per le statistiche descrittive sono:

lo scatter plot: le coppie di campioni (xi, yi) sono plottate in due dimensioni. È una buona modalità per individuare eventuali dipendenze tra variabili. Dall'analisi di questo grafico emerge quanto i dati sono sparsi o concentrati

e se esiste una tendenza lineare nei valori. In genere, se le variabili sono fortemente relazionate, i data point hanno una forma sistematica (una linea, una curva).

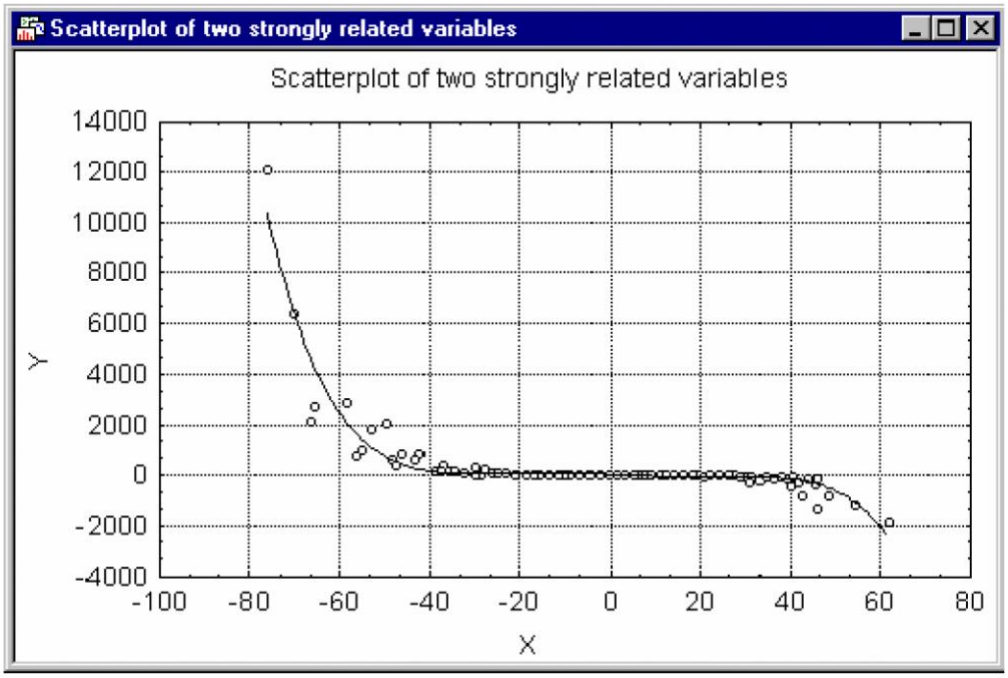

Figura 4.5 – Scatterplot

Il *box plot* è utile per rappresentare dispersione dei campioni di dati. Identifica graficamente i percentili (Figura 4.6). Box plot possono essere costruiti rispetto ai valori medi o alla mediana del campione. Una misura della variabilità basata sui percentili è fornita dalla differenza interquartilica (o differenza interquartile) (Q3-Q1). Poiché tra Q1e Q3 si trova il 50% centrale della distribuzione, se la loro differenza è piccola, vuol dire che la variabilità è contenuta; se la differenza è ampia, la variabilità è elevata. Una rappresentazione grafica è il box plot che è basato sui quartili e sul campo di variazione.

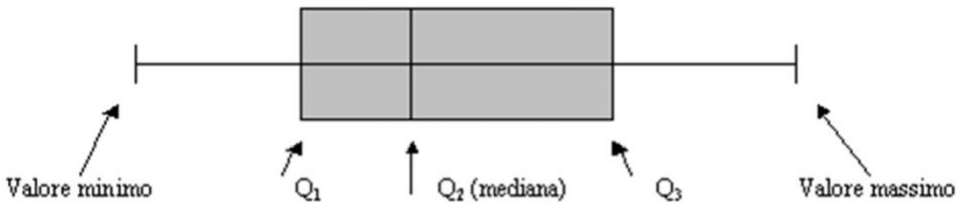

Figura 4.6 – Box plot

Le due linee esterne al rettangolo (box) che si estendono fino al valore più elevato (in alto) e fino al valore più piccolo (in basso) si chiamano baffi (whiskers) e ci dicono se c'è una sostanziale simmetria fra il valore minimo e quello massimo della distribuzione, rispetto al 50 per cento centrale dei valori, che è compreso fra il primo e il terzo quartile.

L'*istogramma* è comunemente utilizzato per rappresentare graficamente serie di dati raggruppati in classi. Fissato un sistema di assi cartesiani ortogonali, sull'asse delle ascisse si riportano tanti intervalli consecutivi quante sono le classi; su questo intervallo si costruiscono dei rettangoli le cui aree sono proporzionali alle frequenze. Si noti che se le classi hanno la stessa ampiezza, o modulo, basterà riportare altezze proporzionali alle frequenze; in caso contrario le altezze si ottengono dividendo la relativa frequenza per l'ampiezza della classe, in modo che l'area rappresenti la frequenza. Con gli istogrammi la somma delle aree di tutti i rettangoli è proporzionale alla somma delle frequenze.

Anno	Maschi	Femmine
1998	28	34
1999	29	30
2000	33	31
2001	35	25
2002	30	27
2003	39	27
2004	45	30

Il grafico con le frequenze relative è, di fatto, lo stesso.

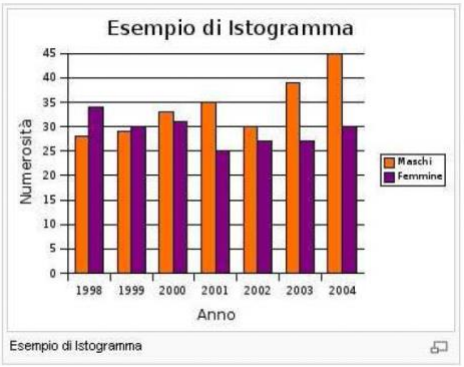

Figura 4.7 – Istogrammi

4.14 Test delle ipotesi

Il test delle ipotesi mira a validare le ipotesi sperimentali formulate durante la pianificazione. I test eseguiti possono essere test parametrici o non parametrici. I test parametrici si basano su un modello che richiede una specifica distribuzione dei dati. Generalmente si assume che i dati abbiano, per fare un test parametrico, una distribuzione normale. L'uso di test parametrici implica che i dati siano misurati almeno su una scala intervallo.

In caso contrario, un test parametrico non può essere utilizzato, e bisognerò optare per un test non parametrico.

Test non-parametrici fanno assunzioni diverse rispetto a quelli parametrici. In particolare, sono più generali rispetto ai test parametrici.

Nel caso di design sperimentale, di un fattore-due trattamenti, il test parametrico da utilizzare è il t-test. Qualora i dati non dovessero soddisfare le condizioni di applicabilità del t-test, sarà utilizzato il rispettivo non-parametrico (Mann-Whitney U test). In particolare, Il t-test è un test parametrico usato per confrontare due campioni indipendenti, usato in progetti con un fattore e due livelli di trattamento.

Di seguito, per il t-test si riporta: input, indica il tipo di dati necessari affinché il test risulti applicato correttamente e le caratteristiche che deve avere l'esperimento da analizzare; H_0, enuncia la ipotesi nulla nell'accezione adatta al test; calcoli, descrive brevemente la procedura di calcolo per lo svolgimento del test; interpretazione, indica come interpretare i risultati della procedura di calcolo in termini di accettazione o rigetto della H_0.

Input	Due campioni indipendenti: $x_1, x_2,\ldots,x_n$ e $y_1, y_2, \ldots, y_m$
H_0	$\mu_x = \mu_y$ ossia, i valori medi previsti sono gli stessi
Calcoli	Calcolare $t_0 = \frac{\bar{x} - \bar{y}}{S_p\sqrt{\frac{1}{n}+\frac{1}{m}}}$ dove: $S_p\sqrt{\frac{(n-1)S_x^2+(m-1)S_y^2}{n+m-2}}$ e dove, S_x^2 e S_y^2 sono le singole varianze
Criterio di interp.	**Bidirezionale**: (H_1: μ_x ? μ_y): rigettare H_0 se $\lvert t_0\rvert > t_{\alpha/2,\, n+m-2}$. In questo caso $t_{\alpha,f}$ è la soglia massima di α con *f* gradi di libertà che è pari a: *n+m-2.* Tale distribuzione è contenuta nella tavola T1 riportata di seguito. **Unidirezionale**: (H_1: $\mu_x > \mu_y$): rigettare H_0 se $t_0 > t_{\alpha,\, n+m-2}$

Tavola T1

Per chiarezza, si fornisce un esempio: È stata confrontata la densità di difetti di programmi diversi in due progetti. I due campioni indipendenti sono: x={3.42, 2.71, 2.84, 1.85, 3.22, 3.48, 2.68, 4.30, 2.49, 1.54} per il primo progetto e y={3.44, 4.97, 4.76, 4.96, 4.10, 3.05, 4.09, 3.69, 4.21, 4.40, 3.49} per il secondo.

L'ipotesi nulla (H_0) è che la densità dei difetti è la stessa in entrambi i progetti.

L'ipotesi alternativa (H_1) è che sono diversi.

Dai dati a disposizione si osserva che: n=10; m=11. I valori medi sono: $\overline{X}$ = 2.853 e $\overline{Y}$ = 4.1055

si calcola: $S_x^2 = 0.6506$ $S_y^2 = 0.4112$ $S_p = 0.7243$ $t_0 = -3.96$

I gradi di libertà sono: f = n+m-2 = 10+11-2 =19. Osservando la tavola T1, si ha che: $t_{0.025,19}$ = 2.093.

Poiché $|t_0| > t_{0.025,19}$ è possibile rigettare l'ipotesi nulla con un test bidirezionale con livello 0.05.

Il Mann-Whitney U test è l'alternativa non parametrica del t-test. Pertanto, può essere applicato al posto di quest'ultimo qualora le assunzioni non fossero verificate.

Di seguito, in per il Mann-Whitney test si riporta: input, indica il tipo di dati necessari affinché il test risulti applicato correttamente e le caratteristiche che deve avere l'esperimento da analizzare; H_0, enuncia la ipotesi nulla nell'accezione adatta al test; calcoli, descrive brevemente la procedura di calcolo per lo svolgimento del test; interpretazione, indica come interpretare i risultati della procedura di calcolo in termini di accettazione o rigetto della H_0

Input	Due campioni indipendenti: $x_1, x_2,\ldots,x_n$ e $y_1, y_2, \ldots, y_m$
H_0	I due campioni provengono dalla stessa distribuzione
Calcoli	Effettuare il ranking (ordinamento) di tutti i campioni e calcolare $U = N_A N_B + \frac{N_A(N_A+1)}{2} - T$ e $U^1 = N_A N_B - U$ Dove $N_A = \min(n,m)$ $N_B = \max(n,m)$ *T* è la somma dei ranking (ordinamenti) del campione più piccolo
Criterio di interp.	La tavola T2, riportata di seguito, è usata per rigettare l'ipotesi nulla secondo i calcoli. Rigettare H_0 se min(*U, U*^{1}) <= del valore nella tavola T2

Tavola T2

Per chiarezza, si fornisce un esempio: Nel fare il ranking si considerano tutte le osservazioni di tutti i campioni. Si assegna a ciascuna osservazione il suo rank (valore nell'ordinamento) a partire dal rank più basso (1) assegnato all'osservazione più piccola. In caso di parità (più osservazioni hanno lo stesso valore) si assegna il rank medio a ciascuna osservazione.

Dato lo stesso campione dell'esempio per il t-test, illustrato precedentemente, si ha che: N_A =min(10,11)=10 e N_B=max(10,11)=11.

Il ranking è:

Progetto 1		Progetto 2	
Valori campione 1	rank	Valori campione 2	rank
3.42	9	3.44	10
2.71	5	4.97	21
2.84	6	4.76	19
1.85	2	4.96	20
3.22	8	4.10	15
3.48	11	3.05	7
2.68	4	4.09	14
4.30	17	3.69	13
2.49	3	4.21	16
1.54	1	4.40	18
		3.49	12

In base ai rank individuati si osserva che: T = 66, U = 99, U_1 = 11

Poiché il min(U, U_1) < 26 (valore individuato nella tavola T2), è possibile rigettare l'ipotesi nulla con un test bidirezionale con livello 0,05.

4.15 Presentazione e Confezionamento dei Risultati

È importante documentare una indagine empirica per renderla replicabile ad altri ricercatori, nonché valorizzare i risultati di ricerca. Inoltre, la formalizzazione dei risultati facilita la generalizzazione e l'applicazione in altri contesti, indipendentemente da chi ha formalizzato la conoscenza acquisita durante la esecuzione del progetto stesso.

La formalizzazione della conoscenza è sintetizzata in pacchetti di conoscenza che sono memorizzati in una fabbrica di esperienza. Questo è realizzato seguendo il modello della Experience Factory illustrato sotto.

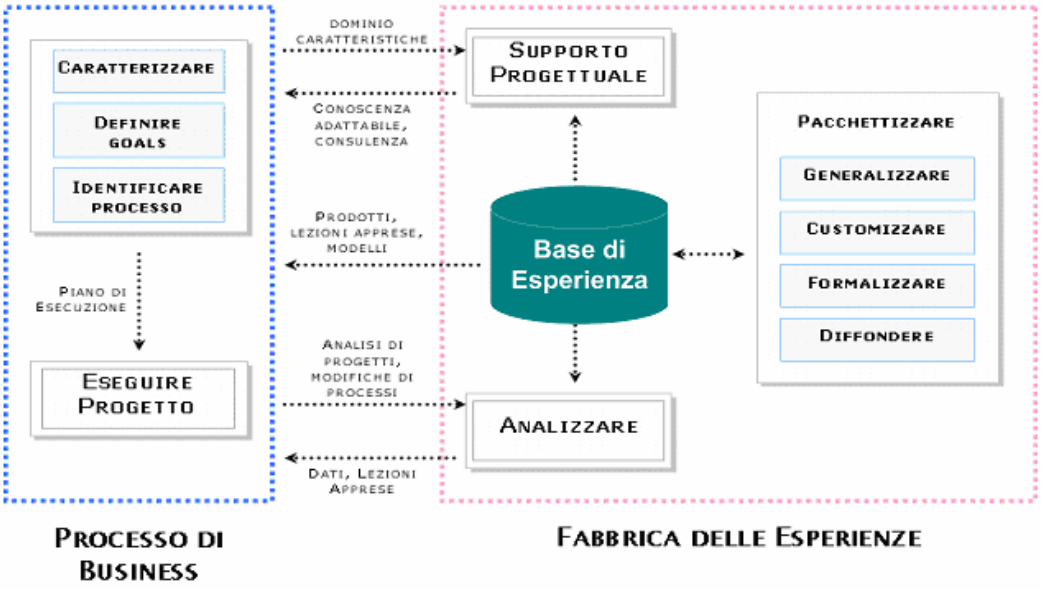

Figura 4.8: Fabbrica delle Esperienze

5. CASO DI STUDIO

5.1 Introduzione al Caso di Studio

Il caso di studio è una metodologia di ricerca adatta per la ricerca di ingegneria del software poiché studia i fenomeni contemporanei nel suo contesto naturale. Tuttavia, la comprensione di ciò che costituisce un caso di studio varia, e quindi la qualità degli studi risultanti. Questo capitolo mira a fornire un'introduzione alla metodologia e alle linee guida del case study per ricercatori che conducono studi di casi e lettori che studiano i rapporti di tali studi. La terminologia e le linee guida sono compilate da diversi manuali di metodologia in altri settori di ricerca, in particolare le scienze sociali e i sistemi di informazione, e adattati alle esigenze nell'ingegneria del software. Presentiamo le pratiche consigliate per i casi di studio nell'ambito della software engineering insieme alle checklist empiricamente validate che ricercatori e professionisti potranno adottare nella progettazione, esecuzione e descrizione di case study. Tali linee guida sono state seguite nella pianificazione ed esecuzione degli studi di validazione realizzate nell'ambito del presente progetto di ricerca.

Il termine "case study" appare di tanto in tanto nel titolo di ingegneria del software documenti di ricerca. Tuttavia, gli studi presentati vanno da molto ambiziosi e bene studi organizzati sul campo, a piccoli esempi di giocattoli che affermano di essere casi studio. Inoltre, ci sono diverse tassonomie utilizzate per classificare la ricerca. Il termine case study è usato in parallelo con termini come studio sul campo e studio osservazionale, ognuno incentrato su a aspetto particolare della metodologia di ricerca. Ad esempio, [Lethbridge, Sim et al.] campo di utilizzo studia come il termine più generale, mentre [Easterbrook, Singer et al.] chiamare case study una delle cinque "classi di metodi di ricerca". Zelkowitz e Wallace propongono una terminologia che è in qualche modo diversa da quella usata in altri campi e categorizzare monitoraggio del progetto, studio di casi e studio sul campo come metodi osservativi [Zelkowitz, Wallace]. Questa pletora di termini provoca confusione e problemi nel tentativo di farlo aggregare studi empirici multipli.

La metodologia del case study è adatta a molti tipi di ingegneria del software la ricerca, poiché gli oggetti di studio sono fenomeni contemporanei, in cui è difficile studiare solitudine. I case study non generano gli stessi risultati ad es. relazioni causali come fanno esperimenti controllati, ma forniscono

una comprensione più profonda dei fenomeni sottostanti studia. Poiché sono diversi dagli studi empirici analitici e controllati, i casi studio sono stati criticati per essere di minor valore, impossibili da generalizzare, di cui essere influenzati ricercatori ecc. Questa critica può essere soddisfatta applicando adeguate pratiche di metodologia di ricerca oltre a riconsiderare che la conoscenza è più che un significato statistico [Flyvbjerg][Lee]. Tuttavia, la comunità di ricerca deve saperne di più sul case study metodologia per esaminarla e giudicarla correttamente. I manuali di metodologia del case study sono superflui disponibili ad es. Scienze sociali [Robson][Stake][Yin] la cui letteratura è stata utilizzata anche nel software ingegneria. Nel campo della ricerca sui sistemi di informazione (IS), la metodologia del case study è anche molto più maturo rispetto all'ingegneria del software. Ad esempio, [Benbasat, Goldstein et al.] fornire una breve panoramica della ricerca di casi studio nei sistemi di informazione, [Lee] analizza casi di studio da una prospettiva positivistica e [Klein, Myers] fa lo stesso da una prospettiva interpretativa.

Lo scopo di questo documento è di fornire una guida per il caso di conduzione del ricercatore studi, per revisori di manoscritti di casi studio e per lettori di documenti di casi studio. È sintetizzato dai manuali di metodologia generale, principalmente dal campo delle scienze sociali, come così come la letteratura nel campo dei sistemi di informazione e adattata all'ingegneria del software esigenze. Naturalmente è inclusa anche la letteratura esistente sugli studi di casi di ingegneria del software. L'analisi sottostante viene effettuata strutturando le informazioni secondo un caso generale studiare il processo di ricerca. Dove diverse raccomandazioni o compaiono i termini, quelli considerati i più adatti al dominio dell'ingegneria del software selezionato, in base all'esperienza degli autori nel condurre studi di casi e leggere casi rapporti di studio. I collegamenti alle fonti di dati sono forniti da riferimenti regolari. In particolare, liste di controllo per ricercatori e lettori si ricava un'analisi sistematica delle liste di controllo esistenti [Höst , Runeson], e successivamente valutato da dottorandi e da membri di l'International Software Engineering Research Network e aggiornato di conseguenza. Questo documento non fornisce dichiarazioni assolute per quello che è considerato un caso "buono" studiare in ingegneria del software. Piuttosto si concentra su una serie di problemi che tutti contribuiscono alla qualità della ricerca. Il requisito minimo per ogni problema deve essere valutato nel suo contesto e molto probabilmente si evolverà nel tempo. Questo è simile ai principi di [Klein, Myers] per casi

studio IS, "dipende dagli autori, revisori ed esercitano il loro giudizio e discrezione nel decidere se, come e quale dei principi dovrebbero essere applicati ". Non valutiamo né lo stato attuale del case study ricerca in ingegneria del software. Questo vale da solo uno studio, simile al sistematico recensione sugli esperimenti di [Sjøberg, Hannay et al.]. Inoltre, entrambi gli esempi sono usati per illustrare buone pratiche e loro mancanza.

5.1.1 Caratteristiche delle Metodologie di Ricerca

Diverse metodologie di ricerca hanno scopi diversi; un tipo di ricerca la metodologia non si adatta a tutti gli scopi. Distinguiamo tra quattro tipi di scopi per ricerca basata sulla classificazione di [Robson]: Ed esplorativo: scoprire cosa sta succedendo, cercare nuove intuizioni e generare idee e ipotesi per nuove ricerche.

- Descrittivo: raffigurante una situazione o un fenomeno
- Esplicativo: cercare una spiegazione di una situazione o di un problema, principalmente ma non necessario sotto forma di relazione causale
- Miglioramento: cercare di migliorare un certo aspetto del fenomeno studiato

La metodologia del caso studio è stata originariamente utilizzata principalmente a fini esplorativi e alcuni ricercatori limitano ancora i casi di studio per questo scopo, come discusso da [Flyvbjerg]. Tuttavia, viene anche utilizzato a scopi descrittivi, se la generalità della situazione o il fenomeno è di secondaria importanza. Casi di studio possono essere utilizzati a scopo esplicativo scopi, ad es. nella progettazione di serie temporali interrotte (studi pre e post evento) sebbene l'isolamento dei fattori può essere un problema. Ciò comporta il test delle teorie esistenti in studi di conferma. Infine, come indicato sopra, case study nell'ingegneria del software la disciplina adotta spesso un approccio di miglioramento, simile alla ricerca d'azione; vedi ad es. il QA studio [Andersson, Runeson 07b]. [Klein, Myers] definiscono tre tipi di casi studio a seconda della prospettiva di ricerca, positivista, critico e interpretativo. Ricerche di un caso positivista prove per proposizioni formali, misura variabili, verifica ipotesi e trae inferenze da un campione a una popolazione dichiarata, cioè è vicino al modello di ricerca in scienze naturali [Lee] e relativi alla categoria esplicativa di Robson. Un caso di studio critico mira al sociale critica e ad essere emancipatori, cioè

identificare diverse forme di sociale, culturale e dominio politico che può ostacolare l'abilità umana. Il miglioramento di casi di studio può avere a carattere di essere critico. Un caso di studio interpretativo tenta di comprendere i fenomeni attraverso l'interpretazione dei partecipanti del loro contesto, che è simile a quella di Robson tipi esplorativi e descrittivi. I casi studio di ingegneria del software tendono a propendere per a prospettiva positivista, specialmente per studi di tipo esplicativo. Lo svolgimento di ricerche su problemi del mondo reale implica un compromesso tra il livello di controllo e grado di realismo. La situazione realistica è spesso complessa e non deterministica, che ostacola la comprensione di ciò che sta accadendo, soprattutto per gli studi esplicativi. D'altro canto, aumentare il controllo riduce il grado di realismo, a volte porta a stabilire i reali fattori influenti al di fuori dell'ambito dello studio. I casi studio sono per definizione condotti in contesti del mondo reale e quindi hanno un alto grado di realismo, principalmente a scapito del livello di controllo. I dati raccolti in uno studio empirico possono essere quantitativi o qualitativi. Quantitativo i dati coinvolgono numeri e classi, mentre i dati qualitativi riguardano parole, descrizioni, immagini, diagrammi ecc. I dati quantitativi vengono analizzati mediante statistiche, mentre i dati qualitativi sono analizzati utilizzando la categorizzazione e l'ordinamento. I casi di studio tendono principalmente a essere qualitativi dati, in quanto forniscono una descrizione più ricca e approfondita. Tuttavia, una combinazione di qualità e i dati quantitativi spesso forniscono una migliore comprensione del fenomeno studiato [Seaman], ovvero quello che a volte viene chiamato "metodi misti" [Robson]. Il processo di ricerca può essere definito fisso o flessibile secondo [Anastas, MacDonald] e [Robson]. In un processo di progettazione fisso, tutti i parametri sono definiti in l'avvio dello studio, mentre in un processo di progettazione flessibile possono essere parametri chiave dello studio cambiato nel corso dello studio. I casi studio sono in genere studi di progettazione flessibili, mentre esperimenti e sondaggi sono studi di progettazione fissi. Altra letteratura usa i termini quantitativi e studi di progettazione qualitativa, rispettivamente per studi di progettazione fissi e flessibili. Preferiamo aderire alla terminologia fissa / flessibile poiché riduce il rischio di confusione con uno studio il design qualitativo può raccogliere dati sia qualitativi che quantitativi. Altrimenti potrebbe essere non è chiaro se il termine qualitativo si riferisca ai dati o alla progettazione dello studio. La triangolazione è importante per aumentare la precisione della ricerca empirica. Triangolazione significa prendere diverse angolazioni rispetto all'oggetto studiato e quindi fornire un più ampia immagine. La necessità di

triangolazione è evidente quando si basa principalmente su dati qualitativi, che è più ampio e ricco, ma meno preciso dei dati quantitativi. Tuttavia, è rilevante anche per dati quantitativi, ad es. per compensare errori di misurazione o modellazione. Quattro possono essere applicati diversi tipi di triangolazione [Dzwigol]:

- Triangolazione dei dati (sorgente): utilizzo di più di un'origine dati o raccolta degli stessi dati in diverse occasioni.
- Triangolazione di osservatori: utilizzando più di un osservatore nello studio.
- Triangolazione metodologica: combinando diversi tipi di metodi di raccolta dei dati, per esempio. metodi qualitativi e quantitativi.
- Triangolazione della teoria — usando teorie o punti di vista alternativi.

Yin aggiunge specificamente alle caratteristiche di un caso di studio che [Yin]:

- “Affronta la situazione tecnicamente distintiva in cui ce ne saranno molti di più variabili rispetto ai punti dati e come risultato
- si basa su più fonti di evidenza, con i dati che devono convergere in una triangolazione moda e come altro risultato
- beneficia del precedente sviluppo di proposizioni teoriche per guidare la raccolta dei dati e analisi ".

In sintesi, le caratteristiche chiave di un caso di studio sono che 1) è di tipo flessibile, per far fronte le caratteristiche complesse e dinamiche dei fenomeni del mondo reale, come l'ingegneria del software, 2) le sue conclusioni si basano su una chiara catena di prove, sia qualitative che quantitative, raccolti da più fonti in modo pianificato e coerente e 3) aggiunge a quelli esistenti la conoscenza basandosi su una teoria precedentemente stabilita, se esiste, o costruendo una teoria.

5.1.2 Processo per eseguire un caso di studio

Nel condurre un caso di studio, ci sono cinque fasi principali del processo da seguire:

1. Progettazione del case study: gli obiettivi sono definiti e il case study è pianificato.

2. Preparazione per la raccolta dei dati: sono definiti procedure e protocolli per la raccolta dei dati.

3. Raccolta delle evidenze: esecuzione con raccolta dei dati sul caso studiato.

4. Analisi dei dati raccolti

5. Report dei risultati

Questo processo è quasi lo stesso per qualsiasi tipo di studio empirico; confrontare ad es. al i processi proposti da [Wohlin, Höst et al. 00] e [Kitchenham, Pfleeger et al]. Tuttavia, a seconda dei casi la metodologia di studio è una strategia di progettazione flessibile, c'è una quantità significativa di iterazione sopra i passaggi [Andersson, Runeson 07 b]. La raccolta e l'analisi dei dati possono essere condotto in modo incrementale. Se vengono raccolti dati insufficienti per l'analisi, più dati la raccolta può essere pianificata ecc. Tuttavia, esiste un limite alla flessibilità; il caso di studio dovrebbe avere obiettivi specifici stabiliti fin dall'inizio. Se gli obiettivi cambiano, è un nuovo caso di studio piuttosto che un cambiamento a quello esistente.

5.1.3 Terminologia

In questo documento, utilizziamo la seguente terminologia. L'obiettivo generale è una dichiarazione di cosa dovrebbe essere raggiunto nel caso studio. Altri possono usare obiettivi, o scopi come sinonimi per obiettivo. L'obiettivo è perfezionato in una serie di ricerche domande alle quali si deve rispondere attraverso l'analisi del case study. Un caso può essere basato su una teoria dell'ingegneria del software. È oltre lo scopo di questo articolo discutere in dettaglio cosa si intende per teoria. Tuttavia, Sjøberg et al. Descrivono un quadro per le teorie compresi i costrutti di interesse, le relazioni tra i costrutti, le spiegazioni alle relazioni, e portata della teoria [Sjøberg, Dybå et al.]. Con questo modo di descrivere teorie, software le teorie di ingegneria comprendono almeno un costrutto dall'ingegneria del software. Una ricerca la domanda può essere correlata a un'ipotesi (a volte chiamata proposizione [Yin]), ovvero a presunta spiegazione di un aspetto del fenomeno in studio. Ipotesi possono in alternativa essere generato dal case study per ulteriori ricerche. Il caso è indicato come l'oggetto dello studio (ad es. un progetto) e contiene una o più unità di analisi (ad es. sottoprogetti). I dati vengono raccolti dai soggetti dello studio, vale a dire quelli che forniscono l'informazione. I dati possono essere quantitativi (numeri, misurazioni) o qualitativi (parole, descrizioni). Un protocollo di case study definisce le procedure dettagliate per la raccolta e analisi dei dati grezzi, a volte chiamati procedure sul campo.

Le linee guida per lo svolgimento di casi studio presentate di seguito sono organizzate secondo questo processo. Le prossime tre sezioni sono dedicate rispettivamente a (i) la definizione degli obiettivi per il case study e la preparazione dei dati raccolta, (ii) la raccolta di dati, (iii) l'analisi dei dati e (iv) linee guida per il reporting dei risultati.

5.2 Progettazione e Pianificazione di un Caso di Studio

5.2.1 Definire il caso

Un piano per un caso di studio dovrebbe contenere almeno i seguenti elementi [Robson]:

- obiettivo: cosa raggiungere?
- Il caso: cosa viene studiato?
- Teoria: quadro di riferimento
- domande di ricerca: cosa sapere?
- metodi: come raccogliere i dati?
- strategia di selezione: dove cercare i dati?

L'obiettivo dello studio può essere, ad esempio, esplorativo, descrittivo, esplicativo o miglioramento. L'obiettivo è naturalmente più generalmente formulato e meno preciso che in progetti di ricerca fissi. L'obiettivo inizialmente è più simile a un punto focale che si evolve durante lo studio. Le domande di ricerca affermano cosa è necessario sapere per soddisfare l'obiettivo dello studio. Simile all'obiettivo, le domande di ricerca si evolvono durante lo studiare e sono limitati a specifiche domande di ricerca durante le iterazioni di studio [Andersson, Runeson 07 b]. Il caso in generale può essere praticamente qualsiasi cosa che sia un "fenomeno contemporaneo in il suo contesto di vita reale" [Yin]. Nell'ingegneria del software, il caso può essere un software progetto di sviluppo, che è la scelta più semplice. In alternativa può essere un individuo, un gruppo di persone, un processo, un prodotto, una politica, un ruolo nell'organizzazione, un evento, una tecnologia, ecc. Il progetto, individuo, gruppo ecc. possono anche costituire un'unità di analisi all'interno di un caso. Nel campo dei sistemi di informazione, il caso può essere "persone fisiche, gruppi ... o un'intera organizzazione. In alternativa, l'unità di analisi può essere specifica progetto o decisione "[Benbasat, Goldstein et al.]. Studi su "programmi giocattolo" o simili sono di ovviamente escluso a causa della sua mancanza di contesto di vita reale. [Yin] distingue tra case study olistici, in cui il caso è studiato nel

suo complesso, e case study integrati dove più unità di analisi sono studiate all'interno di un caso. Se definire a lo studio costituito da due casi olistici o integrati dipende da ciò che definiamo come contesto e obiettivi di ricerca. Nel nostro esempio XP, due progetti sono studiati in due diversi aziende in due diversi domini applicativi, entrambi che utilizzano pratiche agili [Karlström, Runeson 06]. I progetti possono essere considerati due unità di analisi in un caso incorporato studiare se il contesto è quello delle società di software in generale e l'obiettivo della ricerca è studiare agili pratiche. Al contrario, se il contesto è considerato come la società specifica o dominio dell'applicazione, devono essere visti come due casi olistici separati. Benbasat et al. commentare un caso di studio specifico, "Anche se questo studio sembra essere un caso singolo, analisi di unità incorporata, potrebbe essere considerata una progettazione a più casi, a causa della natura centralizzata dei siti. " [Benbasat, Goldstein et al]. In casi di studio, il caso e le unità di analisi dovrebbero essere selezionati intenzionalmente. Questo è in contrasto con sondaggi ed esperimenti, in cui i soggetti vengono campionati da una popolazione a i cui risultati devono essere generalizzati. La selezione dei casi è particolarmente importante quando si replicano casi studio. Un caso di studio può essere letteralmente replicato, ovvero il caso è selezionato per prevedere risultati simili, oppure è teoricamente replicato, ovvero il caso è selezionato per prevedere risultati contrastanti per motivi prevedibili [Yin].

5.2.2 Protocollo del Caso di Studio

Il protocollo del case study è un contenitore per le decisioni di progettazione sul case study e procedure sul campo per la sua realizzazione. Il protocollo è un documento in continuo cambiamento che viene aggiornato quando vengono modificati i piani per il case study. Esistono diversi motivi per mantenere una versione aggiornata di un protocollo di case study. In primo luogo, funge da guida durante la raccolta dei dati e in tal modo impedisce il ricercatore è scomparso per raccogliere i dati che erano previsti per essere raccolti. In secondo luogo, i processi di formulazione del protocollo rendono concreta la ricerca in fase di pianificazione, che può aiutare il ricercatore a decidere quali fonti di dati utilizzare e quali domande porre. In terzo luogo, altri ricercatori e persone rilevanti possono esaminarlo al fine di fornire un feedback i programmi. Il feedback sul protocollo da parte di altri ricercatori può, ad esempio, ridurre il rischio di fonti di dati pertinenti mancanti, domande o ruoli di intervista da includere nella ricerca e assicurare la

relazione tra domande di ricerca e domande di intervista. Infine, può servire come un registro o un diario in cui sono registrati tutti i dati raccolti raccolta e analisi cambiare le decisioni in base alla natura flessibile della ricerca. Questa può essere una fonte importante di informazioni quando viene riportato il caso di studio in un secondo momento. Per tenere traccia delle modifiche durante il progetto di ricerca, il protocollo dovrebbe essere mantenuto sotto qualche forma di controllo della versione. [Pervan, Maimbo] propongono uno schema di un protocollo di studio di caso, che è riassunto in tabella 5.1. Come mostra la proposta, il protocollo è abbastanza dettagliato per supportare una struttura ben strutturata approccio di ricerca.

Section	Content
Preamble	Contains information about the purpose of the protocol, guidelines for data and document storage, publication.
General	Provides a brief overview of the research project and the case research method
Procedures	Detailed description of the procedures for conducting each case, including down-to-earth details on contacts and timing.
Research instrument(s)	Interview guides, questionnaires etc. to be used to ensure consistent data collection.
Data analysis guidelines	Detailed description of data analysis procedures, including data schemas, priori codes etc.
Appendix A	Template letter to invite participants.

Tabella 5.1: Protocollo di un caso di studio secondo Pervan e Maimbo

5.2.3 Checklist

Gli elementi della checklist per la progettazione del case study sono riportati nella Tabella 5.2.

1. What is the case and its units of analysis?
2. Are clear objectives, preliminary research questions, hypotheses (if any) defined in advance?
3. Is the theoretical basis—relation to existing literature or other cases—defined?
4. Are the authors' intentions with the research made clear?
5. Is the case adequately defined (size, domain, process, subjects…)?
6. Is a cause–effect relation under study? If yes, is it possible to distinguish the cause from other factors using the proposed design?
7. Does the design involve data from multiple sources (data triangulation), using multiple methods (method triangulation)?
8. Is there a rationale behind the selection of subjects, roles, artifacts, viewpoints, etc.?
9. Is the specified case relevant to validly address the research questions (construct validity)?
10. Is the integrity of individuals/organizations taken into account?

Tabella 5.2: Checklist per la progettazione di un case study

5.3 Raccolta dei Dati

5.3.1 Fonti per la raccolta dei dati

Esistono diverse fonti di informazioni che possono essere utilizzate in un caso di studio. È importante utilizzare diverse fonti di dati in un caso di studio al fine di limitare gli effetti di un'interpretazione di una singola fonte di dati. Se si può trarre la stessa conclusione da più fonti di informazione, ad es. triangolazione, questa conclusione è più forte di una conclusione basata su un'unica fonte. In un caso di studio è anche importante tenere conto dei punti di vista di diversi ruoli e studiare le differenze, ad esempio, tra diversi progetti e prodotti. Comunemente, le conclusioni sono tratte analizzando le differenze tra le fonti di dati. Secondo [Lethbridge, Sim et al.] le tecniche di raccolta dei dati possono essere suddivise in tre livelli:

- Primo livello: metodi diretti significa che il ricercatore è in contatto diretto con il soggetto e raccogliere dati in tempo reale. Questo è il caso, ad esempio, di interviste, focus group, sondaggi Delphi [Dalkey, Helmer] e osservazioni con “thinking aloud ".
- Secondo livello: metodi indiretti in cui il ricercatore raccoglie direttamente dati grezzi senza interagendo effettivamente con i soggetti durante la raccolta dei dati. Questo approccio è, per esempio preso in Software Project Telemetry [Johnson, Kou et al.] in cui l'utilizzo di gli strumenti di ingegneria del software vengono monitorati automaticamente e osservati attraverso i video registrazione.
- Terzo livello: analisi indipendente di manufatti di lavoro ove già disponibili e a volte vengono utilizzati i dati compilati. Questo è ad esempio il caso di documenti come vengono analizzate le specifiche dei requisiti e le relazioni sugli errori di un'organizzazione o quando vengono analizzati dati da database organizzativi come la contabilità temporale.

I metodi di primo livello sono per lo più costosi da applicare rispetto al secondo o terzo livello, poiché richiedono sforzi significativi sia da parte del ricercatore che dei soggetti. Il vantaggio dei metodi di primo e secondo livello è che il ricercatore può in larga misura controllare esattamente quali dati vengono raccolti, come vengono raccolti, in quale forma vengono raccolti, quale sia il contesto ecc. I metodi di terzo grado sono per lo più meno costosi, ma non lo sono offrire lo stesso controllo al ricercatore;

quindi, la qualità dei dati non è sotto controllo né, né per quanto riguarda la qualità dei dati originali né il suo utilizzo ai fini del case study. In molti casi il ricercatore deve, in una certa misura, basare i dettagli della raccolta dei dati su quali dati sono disponibili. Per i metodi di terzo livello si dovrebbe anche notare che i dati sono stati raccolti e registrati per un altro scopo rispetto a quello dello studio di ricerca, contrariamente alle linee guida generali sulle metriche [van Solingen, Berghout].

5.3.2 Checklist

Gli elementi della checklist per la preparazione ed esecuzione del case study sono riportati nelle Tabelle 5.3 e 5.4.

11. Is a case study protocol for data collection and analysis derived (what, why, how, when)? Are procedures for its update defined?
12. Are multiple data sources and collection methods planned (triangulation)?
13. Are measurement instruments and procedures well defined (measurement definitions, interview questions)?
14. Are the planned methods and measurements sufficient to fulfill the objective of the study?
15. Is the study design approved by a review board, and has informed consent obtained from individuals and organizations?

Tabella 5.3: Checklist per la preparazione della raccolta dei dati di un case study

16. Is data collected according to the case study protocol?
17. Is the observed phenomenon correctly implemented (e.g. to what extent is a design method under study actually used)?
18. Is data recorded to enable further analysis?
19. Are sensitive results identified (for individuals, the organization or the project)?
20. Are the data collection procedures well traceable?
21. Does the collected data provide ability to address the research question?

Tabella 5.4: Checklist per la raccolta di evidenze di un case study

5.4 Analisi dei Dati

5.4.1 Analisi Quantitativa dei Dati

L'analisi dei dati viene condotta in modo diverso per i dati quantitativi e qualitativi. Per dati quantitativi, l'analisi include in genere analisi di statistiche descrittive, analisi di correlazione, sviluppo di modelli predittivi e test di ipotesi. Tutte queste attività sono rilevanti nella ricerca di casi

studio. Statistiche descrittive, come valori medi, deviazioni standard, istogrammi e scatter plot, vengono utilizzati per comprendere i dati raccolti. Correlazioni vengono condotte analisi e sviluppo di modelli predittivi al fine di descrivere come la misurazione da un'attività di processo successiva è correlata a una misurazione di processo precedente. Il test di ipotesi viene condotto al fine di determinare se vi sia un effetto significativo di uno o più variabili (variabili indipendenti) su una o più altre variabili (variabili dipendenti). Va notato che i metodi per l'analisi quantitativa presuppongono un design definito. Ad esempio, se una domanda con una risposta quantitativa viene modificata a metà di una serie delle interviste, ciò rende impossibile interpretare il valore medio delle risposte. Ancora, insiemi di dati quantitativi da singoli casi tendono ad essere molto piccoli, a causa del numero di intervistati o punti di misurazione, che causano particolari preoccupazioni nell'analisi.

5.4.2 Analisi Qualitativa dei dati

Poiché la ricerca di casi studio è un metodo di ricerca flessibile, metodi di analisi dei dati qualitativi [Seaman] sono comunemente usati. L'obiettivo di base dell'analisi è derivare conclusioni dai dati, mantenendo una chiara catena di prove. La catena di prove significa che un lettore dovrebbe essere in grado di seguire la derivazione dei risultati e delle conclusioni dai dati raccolti [Yin]. Ciò significa che informazioni sufficienti da ogni passaggio dello studio e ogni decisione presa dal ricercatore deve essere presentata. Oltre alla necessità di tenere a mente una chiara catena di prove, nell'analisi qualitativa la ricerca si caratterizza per l'esecuzione di analisi in parallelo con la raccolta dei dati e la necessità di tecniche di analisi sistematica. L'analisi deve essere eseguita in parallelo con la raccolta dei dati poiché l'approccio è flessibile e vengono scoperti nuovi approfondimenti durante l'analisi. Al fine di indagare su questi approfondimenti, spesso devono essere raccolti nuovi dati, e la strumentazione come i questionari o le interviste devono essere aggiornate. La necessità di essere sistematico è il risultato diretto del fatto che le tecniche di raccolta dei dati possono essere costantemente aggiornate, mentre allo stesso tempo viene richiesto di mantenere una catena di prove. Al fine di ridurre la distorsione da parte dei singoli ricercatori, l'analisi trae vantaggio dall'essere condotto da più ricercatori. I risultati preliminari di ogni singolo ricercatore vengono uniti in un risultato di analisi comune in un secondo passaggio. Tenere traccia e

segnalare lo schema di cooperazione aiuta ad aumentare la validità dello studio e delle conclusioni.

5.4.3 Validità dei risultati

La validità di uno studio denota l'affidabilità dei risultati, in che misura i risultati sono veri e non distorti dal punto di vista soggettivo dei ricercatori. Ovviamente è troppo tardi per considerare la validità durante l'analisi. La validità deve essere affrontata durante tutte le fasi precedenti del case study. Esistono diversi modi per classificare gli aspetti della validità e le minacce alla validità in letteratura. Abbiamo scelto uno schema di classificazione che viene utilizzato anche da [Yin] e simile a quello che viene solitamente utilizzato negli esperimenti controllati nell'ingegneria del software [Wohlin, Höst et al. 00]. Alcuni ricercatori hanno sostenuto di avere un diverso schema di classificazione per studi di progettazione flessibili (credibilità, trasferibilità, affidabilità, confermabilità), mentre noi preferiamo rendere operativo questo schema per studi di progettazione flessibili, invece di cambiare i termini. Questo schema distingue tra quattro diversi aspetti di validità, che possono essere riassunti come segue:

Validità di Costrutto: questo aspetto della validità riflette in che misura le misure operative studiate rappresentano realmente ciò che il ricercatore ha in mente e ciò che viene investigato in base alle domande di ricerca. Se, ad esempio, i costrutti discussi nelle domande dell'intervista non sono interpretati nello stesso modo dal ricercatore e le persone intervistate, esiste una minaccia alla validità del costrutto.

Validità interna: questo aspetto della validità è preoccupante quando si esaminano le relazioni causali. Quando il ricercatore sta studiando se un fattore influisce su un fattore indagato, esiste il rischio che anche il fattore indagato sia influenzato da un terzo fattore. Se il ricercatore non è a conoscenza del terzo fattore e / o non sa fino a che punto influisce sul fattore indagato, esiste una minaccia alla validità interna.

Validità esterna: questo aspetto della validità riguarda la misura in cui è possibile generalizzare i risultati e in che misura i risultati sono di interesse per altre persone al di fuori del caso indagato. Durante l'analisi della validità esterna, il ricercatore cerca di analizzare in che misura i risultati sono rilevanti per altri casi. Non c'è popolazione dalla quale è stato estratto un campione statisticamente rappresentativo. Tuttavia, per casi di studio, l'intenzione è quella di consentire la generalizzazione analitica in cui i

risultati sono estesi a casi che hanno caratteristiche comuni e quindi per i quali i risultati sono rilevanti, vale a dire definire una teoria.

Affidabilità: questo aspetto riguarda la misura in cui i dati e l'analisi dipendono dai ricercatori specifici. Ipoteticamente, se in seguito un altro ricercatore ha condotto lo stesso studio, il risultato dovrebbe essere lo stesso. Le minacce a questo aspetto della validità sono, ad esempio, se non è chiaro come codificare i dati raccolti o se i questionari o le domande dell'intervista non sono chiare.

Come preannunciato sopra, è importante considerare la validità del case study fin dall'inizio. Esempi di modi per migliorare la validità sono la triangolazione, lo sviluppo e il mantenimento di un protocollo dettagliato del case study, far revisionare i progetti, protocolli, ecc. da altri ricercatori, aver raccolto dati e ottenuto risultati che sono revisionati da altri soggetti, aver posto particolare attenzione per l'analisi dei casi negativi.

5.4.4 Checklist

Gli elementi della checklist per l'analisi dei dati di un case study sono riportati nella Tabella 5.5.

22. Is the analysis methodology defined, including roles and review procedures?
23. Is a chain of evidence shown with traceable inferences from data to research questions and existing theory?
24. Are alternative perspectives and explanations used in the analysis?
25. Is a cause–effect relation under study? If yes, is it possible to distinguish the cause from other factors in the analysis?
26. Are there clear conclusions from the analysis, including recommendations for practice/further research?
27. Are threats to the validity analyzed in a systematic way and countermeasures taken? (Construct, internal, external, reliability)

Tabella 5.5: Checklist per l'analisi dei dati di un case study

5.5 Report dei Risultati

Uno studio empirico non può essere distinto dal suo resoconto. Il report dei risultati comunica i risultati dello studio, ma è anche la principale fonte di informazioni per giudicare la qualità dello studio stesso. I report possono avere un pubblico diverso, come ricercatori, responsabili delle politiche, sponsor della ricerca e professionisti del settore [Yin]. Ciò può comportare la necessità di scrivere report diversi per un pubblico diverso. Di seguito, discutiamo prima delle caratteristiche di un report di un caso di studio e poi di una struttura proposta.

5.5.1 Caratteristiche

Robson definisce un insieme di caratteristiche che dovrebbe avere una relazione di un caso di studio [Robson], che in sintesi implica che dovrebbe:

- specificare di cosa parla lo studio
- comunicare un chiaro senso del caso studiato
- fornire una "storia dell'indagine" in modo che il lettore possa vedere cosa è stato fatto, da chi e come.
- fornire dati in forma mirata, in modo che il lettore possa assicurarsi che le conclusioni siano ragionevoli
- articolare le conclusioni del ricercatore e inserirle in un contesto che influenzano.

Inoltre, ciò deve avvenire nel rispetto dell'obbligo e della finalità del ricercatore di pubblicare i loro risultati e l'integrità delle aziende e degli individui [Amschler, Pradhan] coinvolti. La segnalazione degli obiettivi del caso di studio e delle domande di ricerca è piuttosto semplice. Se cambiassero sostanzialmente nel corso dello studio, questo dovrebbe essere segnalato per aiutare a comprendere il caso.

Descrivere il caso potrebbe essere più sensibile, poiché ciò potrebbe consentire l'identificazione del caso o dei suoi soggetti. Tuttavia, il caso potrebbe essere meglio caratterizzato da mezzi diversi dal dominio e dal paese dell'applicazione. Le caratteristiche interne, come le dimensioni dell'unità studiata, l'età media del personale, ecc. possono essere più interessanti delle caratteristiche esterne come il dominio e il turnover. O il caso costituisce una piccola sotto unità di una grande società, e quindi difficilmente può essere identificato tra le numerose sotto unità, oppure è una piccola azienda e quindi è difficile identificarla tra molti candidati. Tuttavia, bisogna fare attenzione a trovare questo equilibrio.

I dati sono raccolti in abbondanza in uno studio qualitativo e l'analisi ha come obiettivo principale la riduzione e l'organizzazione dei dati per fornire una catena di prove per le conclusioni. Tuttavia, per stabilire la fiducia nello studio, il lettore necessita di istantanee pertinenti dai dati che supportano le conclusioni. Queste istantanee possono essere ad es. sotto forma di citazioni (dichiarazioni tipiche o speciali), immagini o narrazioni con soggetti anonimizzati. Inoltre, le categorie utilizzate nella classificazione dei dati, che portano a determinate conclusioni, possono aiutare il lettore a seguire la catena di prove. Infine, le conclusioni devono essere riportate e inserite in un contesto di implicazioni, ad es. formando teorie. Un caso di studio non

può essere generalizzato nel senso di essere rappresentativo di una popolazione, ma questo non è l'unico modo per raggiungere e trasferire conoscenze. Le conclusioni possono essere tratte senza statistiche e possono essere interpretate e correlate ad altri casi.

5.5.2 Struttura

Yin propone diverse strutture alternative per la segnalazione di casi studio in generale [Yin].

- Analitica lineare: la struttura standard del rapporto di ricerca (problema, lavoro correlato, metodi, analisi, conclusioni)
- Comparativo: lo stesso caso viene ripetuto due o più volte per confrontare descrizioni, spiegazioni o punti di vista alternativi.
- Cronologico: una struttura più adatta per studi longitudinali.
- Theory-building — presenta il caso secondo una logica di costruzione della teoria al fine di costituire una catena di prove per una teoria.
- Suspense — ripristina la struttura lineare-analitica e riporta prima le conclusioni e poi le supporta con prove.
- Unsequenced: nessuna delle precedenti, ad es. quando si segnalano le caratteristiche generali di una serie di casi.

Per il contesto dei casi studio su cui ci concentriamo, la struttura lineare-analitica è la struttura più adatta. La struttura di alto livello per la segnalazione di esperimenti di ingegneria del software proposta da [Jedlitschka, Pfahl] si adatta quindi anche allo scopo di come relazionare e sintetizzare i case study. Tuttavia, sono necessari alcuni cambiamenti, basati su caratteristiche specifiche di casi studio e altri problemi basati su una valutazione condotta da [Kitchenham, Al-Khilidar et al.]. La struttura risultante è presentata nella Tabella 5.6. In un caso di studio, la teoria può costituire un quadro per l'analisi; quindi, ci sono due tipi di lavori correlati: a) studi precedenti sull'argomento e b) teorie su cui si basa l'attuale studio. La sezione di progettazione corrisponde al protocollo del caso di studio, ovvero riporta la pianificazione del caso di studio, comprese le misure adottate per garantire la validità dello studio.

Poiché il case study ha una progettazione flessibile e la raccolta e l'analisi dei dati sono più intrecciate, queste sezioni possono essere combinate in una sola. Di conseguenza, i contenuti di livello inferiore devono essere adeguati, come proposto nella Tabella 5.6. In particolare, per la sezione di dati combinati, lo schema di codifica costituisce spesso una struttura

naturale di sottosezione. In alternativa, per uno case study comparativo, la sezione di dati può essere strutturata in base ai casi a confronto e, per uno studio longitudinale, la scala temporale può costituire la struttura della sezione di dati. Questa sezione dei risultati combinati include anche una valutazione della validità dei risultati finali.

Experiments		Case studies	
Section heading	Subsections	Section headings	Subsections
Title		Title	
Authorship		Authorship	
Structured abstract		(Structured) abstract	
Motivation	Problem statement	Introduction	Problem statement
	Research objectives		Research objectives
	Context		Context
Related work		Related work	Earlier studies
			Theory
Experimental design	Goals, hypotheses and variables	Case study design	Research questions
	Design		Case and subjects selection
	Subjects		Data collection procedure(s)
	Objects		Analysis procedure(s)
	Instrumentation		Validity procedure(s)
	Data collection procedure		
	Analysis procedure		
	Evaluation of validity		
Execution	Sample	Results	Case and subjects description, covering execution, analysis and interpretation issues.
	Preparation		Subsections may be structured e.g. according to coding scheme, each linking observations to conclusions
	Data collection performed		Evaluation of validity
	Validity procedure		
Analysis	Descriptive statistics		
	Data set reduction		
	Hypothesis testing		
Interpretation	Evaluation of results and implications		
Conclusions and Future work	Relation to existing evidence	Conclusions and Future work	Summary of conclusions
	Impact		Relation to existing evidence
	Limitations		Impact/Implications
	Future work		Limitations
			Future work
Acknowledgements		Acknowledgements	
References		References	

Tabella 5.6: Struttura per il Report dei Risultati di un Case Study vs la struttura di un Report di un Esperimento Controllato

5.5.3 Checklist

Gli elementi della checklist per elaborare il report dei risultati del case study sono riportati nella Tabella 5.7.

28. Are the case and its units of analysis adequately presented?
29. Are the objective, the research questions and corresponding answers reported?
30. Are related theory and hypotheses clearly reported?
31. Are the data collection procedures presented, with relevant motivation?
32. Is sufficient raw data presented (e.g. real life examples, quotations)?
33. Are the analysis procedures clearly reported?
34. Are threats to validity analyses reported along with countermeasures taken to reduce threats?
35. Are ethical issues reported openly (personal intentions, integrity issues, confidentiality)
36. Does the report contain conclusions, implications for practice and future research?
37. Does the report give a realistic and credible impression?
38. Is the report suitable for its audience, easy to read and well structured?

Tabella 5.7: Checklist per il report dei risultati di un case study

6. RICERCA AZIONE

Scopo di questa sezione è quelli di introdurre la Ricerca – Azione Canonica (CAR) per permettere ai lettori meno esperti di comprendere come i progetti, oggetto di questo lavoro, sono stati formalizzati e condotti. Viene tracciato un breve riepilogo storico sulle cause e necessità che hanno portato alla formalizzazione della CAR e ne vengono elencati i principi. Non è obiettivo di questo lavoro studiare e dettagliare tali principi.

La Ricerca – Azione è stata elogiata per la rilevanza dei suoi risultati, ma è stata anche criticata perché priva di rigore. Di conseguenza, è utile chiarire il significato e l'importanza del rigore e della rilevanza. Esistono diverse definizioni di rigore, la più rilevante per il nostro scopo è "l'uso corretto dei metodi e l'analisi appropriata delle attività". Questa definizione implica l'esistenza di un contesto che consente di determinare cosa è corretto e appropriato.

Inoltre, il rigore deve essere supportato da principi che incorporano una solida base teorica e sono accettati da una comunità di ricerca che supporta la loro applicazione riflessiva e appropriata in contesti problematici. Allo stesso modo, si è osservato che non esiste "nulla di così pratico come una buona teoria". Ciò non confuta la critica secondo cui la Ricerca – Azione non è sufficientemente rigorosa, ma identifica la necessità di linee guida pratiche per incoraggiare il rigore nella Ricerca – Azione.

Come già detto la rilevanza dipende dal contesto, poiché è essenziale identificare qual è l'obiettivo dello studio e verso chi è rivolto. Si sostiene che le ricerche rilevanti si focalizzano sulla pratica e hanno come obiettivo quello di fornire soluzioni reali ai professionisti del settore IS, inoltre, devono essere anche rilevanti per il dominio di ricerca nel quale i ricercatori operano. Questi due fattori, la natura dell'argomento selezionato e le implicazioni dei risultati, possono essere utilizzati per valutare la rilevanza di uno studio.

Rigore e rilevanza sono sempre stati percepiti come due attributi la cui relazione è inversa, questo induce a pensare che o cambiamo i nostri standard, abbandonando il nostro metodo, o preserviamo la nostra integrità intellettuale. Questa prospettiva è "ingenua", poiché produrre una ricerca rigorosa e pertinente è sia desiderabile che applicabile. I due attributi non devono essere visti come mutuamente esclusivi, sebbene risultino difficili da ottenere in un progetto di Ricerca – Azione.

I principi presentati sono stati definiti analizzando la letteratura relativa alla Ricerca – Azione, partendo dalle prime pubblicazioni degli anni 40 nel campo delle scienze sociali. I principi sono specifici per il CAR e non coprono questioni come l'etica della conduzione di una ricerca, indipendentemente dal metodo utilizzato, o dalla posizione epistemologica del ricercatore. In aggiunta ai principi, sono stati individuati dei criteri per supportare il rispetto di ognuno dei principi in uno specifico progetto CAR. Questi criteri possono essere usati dai ricercatori, per pianificare e condurre in progetto, e dai lettori, per valutare il report di tale progetto. Ogni criterio è espresso in forma di quesito del tipo "Qualcosa è stata fatta?". I criteri sono progettati per essere pratici ai ricercatori.

I cinque principi proposti per il CAR sono:

1. Il principio del "Researcher – Client Agreement" (RCA)
2. Il principio del "Cyclical Process Model" (CPM)
3. Il principio della Teoria
4. Il principio del "Change through Action"
5. Il principio del "Learn through Reflection"

6.1 Accordo tra ricercatori e committente (RCA)

RCA è alla base di un progetto AR (Ricerca – Azione). Tuttavia, affinché la RCA sia efficace, è necessario che il cliente comprenda come un progetto CAR funziona e quali sono i vantaggi e gli svantaggi per l'organizzazione. Raggiungere questa comprensione può richiedere un processo di trasferimento della conoscenza, dal ricercatore al cliente. L'accordo dovrebbe contenere garanzie reciproche per il comportamento nel contesto del progetto. Una RCA ben strutturata dovrebbe fornire una solida base per creare fiducia tra i vari stakeholder e contribuire alla validità interna della ricerca. L'accordo aiuta a infondere lo spirito di ricerca condivisa, facendo in modo che i clienti contribuiscano mentre il ricercatore determina gli obiettivi, pianifica le azioni, implementa i cambiamenti e valuta i risultati di tali cambiamenti. L'aderenza a questo principio può essere valutata utilizzando i criteri mostrati in *Tabella* 6.1.

1a	I ricercatori ed il client concordano sul fatto che CAR è l'approccio appropriato per la situazione organizzativa?
1b	L'obiettivo della ricerca è esplicito e specificato chiaramente?
1c	Il cliente ha formalmente richiesto il progetto di ricerca?
1d	I ruoli e le responsabilità dei ricercatori e dei membri dell'organizzazione sono specificati esplicitamente?
1e	Gli obiettivi del provetto e le misure di valutazione sono specificati esplicitamente?
1f	Il modo in cui raccogliere i dati e i metodi di analisi sono specificati esplicitamente?

Tabella 6.1 – Criteri di valutazione RCA

6.2 Modello di processo ciclico

Definito l'accordo, inizia l'azione da parte dei ricercatori per lavorare sul progetto. Le attività seguiranno un modello di processo ciclico (CPM). Il modello proposto è definito facendo tesoro sulle esperienze raccolta dalla letteratura ed è mostrato in *Figura* 6.1.

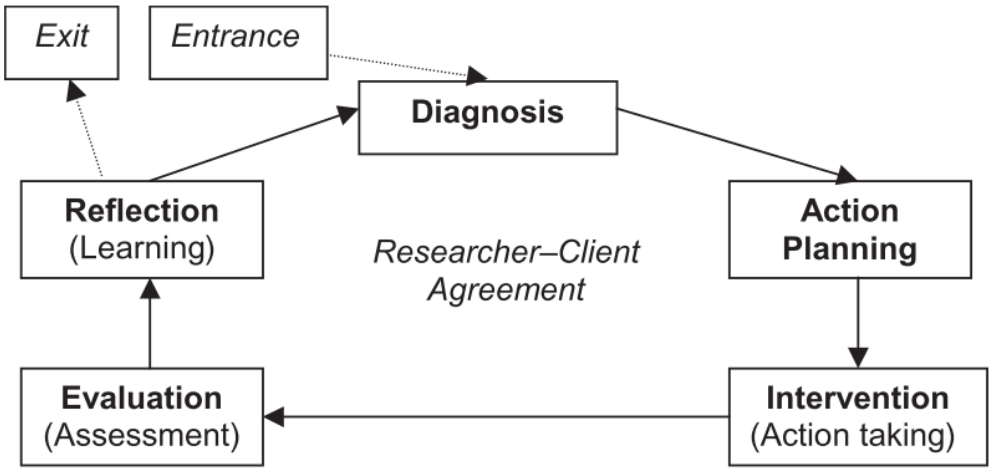

Figura 6.1 – RCA Cyclical Process Model (CPM)

Avanzare attraverso il CPM in modo sequenziale contribuirà a garantire che un progetto CAR sia condotto con rigore sistematico, una caratteristica distintiva della CAR. In alcuni casi è possibile completare un progetto in modo soddisfacente in una singola iterazione, ma spesso sono necessarie più iterazioni.

La misura in cui il principio CPM è seguito in un progetto può essere descritta dall'aderenza ai sette criteri mostrati in *Tabella* .

2a	Il progetto segue il modello CPM? Se no, ogni deviazione è giustificata?
2b	I ricercatori hanno condotto una diagnosi indipendente sulla situazione organizzativa?
2c	Le azioni pianificati son basate esplicitamente sui risultati della diagnosi?
2d	Dove le azioni pianificate sono eseguite e valutate?
2e	Il ricercatore ha effettuato una riflessione sui risultati degli interventi?
2f	La riflessione è seguita da una decisione sul procedere o meno con un'ulteriore iterazione?

2d	L'uscita del ricercatore e la conclusione del progetto sono giustificate in maniera articolata?

Tabella 6.2 – Criteri di valutazione CPM

6.3 Teoria

AR senza una teoria non può considerarsi ricerca. Si insiste sul fatto che un articolato framework teorico deve essere imposto sul fenomeno di interesse. Altri, al contrario, credono che l'applicazione della teoria, in particolare nelle prime fasi del progetto potrebbe essere contro produttiva, infatti, è altamente improbabile che il ricercatore possa conoscere in modo definitivo e in anticipo la teoria esatta che verrà usata o sviluppata. Un progetto CAR può iniziare con l'action learning senza teoria. Tuttavia, similmente al metodo scientifico tradizionale, la fase di diagnostica fornisce un punto di partenza per il confronto per la valutazione post-implementazione. Viene suggerito che se una teoria "radicata" non emerge dalla diagnosi, allora la teoria esplicativa è necessaria come parte della fase di pianificazione.

Nell'ingegneria del Software come nelle scienze sociali il processo per la produzione della teoria deve essere iterativo. Perché, prodotta una teoria, questa può spiegare solo parte di ciò che si vede nella realtà. Ciò nonostante, il processo di sviluppo di una teoria deve essere sempre attivo. perché è meglio avere una parziale comprensione che può servire come piattaforma per futuri miglioramenti della stessa teoria piuttosto che scartare una teoria intermedia solo perché non riesce a spiegare tutto.

Il ruolo essenziale della teoria nei progetti CAR è rappresentato dai criteri mostrati in *Tabella* .

3a	Le attività di progetto sono guidate da una teoria o da un insieme di teorie?
3b	Il dominio investigato, e le specifiche impostazioni del problema, sono rilevanti e significativi per la comunità dei ricercatori tanto quanto lo sono per il cliente?
3c	Viene utilizzato un modello teorico per derivare le cause del problema osservato?
3d	L'intervento pianificato segue il modello teorico?

3e	La valutazione dei risultati e degli interventi è guidata da una teoria?

Tabella 6.3 – Criteri di valutazione della Teoria

6.4 Cambiare attraverso l'azione

L'essenza dei progetti CAR è effettuare azioni al fine di cambiare la situazione corrente e le sue condizioni non soddisfacenti. Il quarto principio rispecchia questa essenza e sottolinea l'indivisibilità dell'azione e del cambiamento. Una mancanza di cambiamenti in condizioni non soddisfacenti suggerisce che non c'è un problema significativo, che l'intervento ha fallito nell'indirizzare i problemi esistenti o che l'attuale situazione non può essere alterata perché una pratica politica ostacola i cambiamenti indotti dal CAR.

I criteri correlati a questo principio sono presenti in *Tabella* 6.4.

4a	I ricercatori ed il cliente sono entrambi motivati a migliorare la situazione?
4b	Il problema e le cause ipotizzare sono specificate come risultato della fase di diagnosi?
4c	Sono state pianificate le azioni per affrontare la causa ipotizzate?
4d	Il cliente ha approvato le azioni pianificate prima della loro implementazione?
4e	La situazione dell'organizzazione è stata valutata in modo completo sia prima che dopo l'intervento?
4f	I tempi e la natura delle azioni sono stati chiaramente e completamente documentati?

Tabella 6.4 – Criteri di valutazione delle azioni

6.5 Apprendere attraverso le riflessioni

L'esplicitazione delle esperienze apprese è l'attività più critica in un progetto AR. Il fondamento logico del principio di apprendimento attraverso la riflessione deriva dalle molteplici responsabilità del ricercatore: verso i clienti e verso la comunità di ricerca. Ciò è coerente con la richiesta comune

di rapporti di ricerca per specificare le implicazioni per la pratica e la ricerca. I clienti si concentreranno sui risultati pratici, mentre la comunità di ricerca sarà interessata alla scoperta di nuove conoscenze. I progressi pratici e il progresso della conoscenza derivano entrambi da riflessioni. I criteri correlati a questo principio sono presenti in Tabella 6.5.

5a	I ricercatori hanno fornito un rapporto contente i progressi ai membri dell'organizzazione?
5b	I ricercatori e il cliente hanno riflettuto riguardo i risultati del progetto?
5c	Le attività svolte ed i relativi risultati sono riportati in maniera chiara e completa?
5d	I risultati sono stati considerati in termini di implicazioni per ulteriori azioni?
5e	I risultati sono stati considerati in termini di implicazioni per l'azione da intraprendere in settori di ricerca correlati?
5f	I risultati sono stati considerati in termini di implicazioni per la comunità di ricerca (conoscenza generale, teoria di informazione)?
5g	I risultati sono stati considerati in termini dell'applicabilità generale di CAR?

Tabella 6.5 – Criteri di valutazione delle riflessioni

7. LE INDAGINI DEL PROGETTO I-MOLE

Dopo aver analizzato i diversi approcci all'esecuzione di un esperimento e le caratteristiche di ognuno di essi, tenendo conto che trattasi di ricerca industriale che ha come obiettivo ultimo quello di ottenere risultati da contesti reali e non realistici, l'approccio di indagine empirica scelto per tutti i quesiti posti è "il caso di studio".

Per la validazione del progetto I-Mole sono state previste indagini che si svilupperanno in 5 aree:

- Manutenzione rapida ed affidabile
- Composizione dei servizi offerti dall'ecosistema I-Mole
- Specificazione dei servizi offerti dall'ecosistema I-Mole
- Ricerca dei Servizi offerti dall'ecosistema I-Mole
- Usabilità e User eXperience (UX).

La progettazione dei casi di studio utili all'indagine è stata effettuata con l'obiettivo di minimizzare i fattori che possono falsare la valutazione dei risultati della stessa.

A tal fine, i casi di studio qui progettati saranno eseguiti su un servizio pilota visto che l'innovazione proposta nel progetto I-Mole ha già un buon livello di maturità.

Pertanto, ogni caso di studio dovrà prevedere l'esecuzione di task da parte dei soggetti sperimentali monitorata da un esperto di Assioma che dovrà misurare le variabili di volta in volta definite e gestire la compilazione dei questionari utili a determinare l'efficacia dell'innovazione nei processi produttivi dell'impresa che sperimenta.

7.1 Manutenzione rapida ed affidabile

L'**obiettivo** dell'indagine è la validazione dell'influenza dell'uso delle tecnologie IH, FR, SPL su:

- La modificabilità della piattaforma
- Il mantenimento della qualità del software nonostante le modifiche.

L'indagine prevederà le seguenti **azioni**

- con la esecuzione delle indagini anzidette e/o con la presentazione della piattaforma e/o con le modifiche richieste dagli sviluppatori per

miglioramenti tecnici, a diversi stadi dello sviluppo della stessa (ovvero con differenti MVP), si rilevano delle modifiche da eseguire.
- Si selezionano tra queste le modifiche che gli sperimentatori ritengono più significative per la struttura e la qualità della piattaforma e con esse si rilevano gli indicatori che possono esprimere il raggiungimento dell'obiettivo.

Le **misure** rilevate:

- **Impegno-persona** speso per la **realizzazione** di ogni modifica effettuata
- **Impegno persona** per la produzione del piano di test e di regressione da effettuare
- **Numero casi di test** totali contenuti nel piano,
- **Numero di casi di test** eseguiti sino alla eliminazione di tutti gli errori;
- **Numero di sessioni di test** eseguiti sino alla eliminazione di tutti gli errori;
- **Technical debt** prima e dopo la modifica

7.2 Composizione dei servizi

L'obiettivo di questa macroarea è quello di validare la complessità di combinare servizi già esistenti per produrre un servizio da proporre a catalogo. Dovranno essere combinati sia servizi primi che servizi compositi da più servizi primi/compositi.

L'indagine empirica prevederà la seguente azione:

- si deve comporre un nuovo servizio che, quindi, non sia nel catalogo dell'ecosistema, utilizzando servizi che sono nel catalogo. Ad ogni soggetto sperimentale si dovrà assegnare un servizio da comporre. Il servizio assegnato a tutti i soggetti sperimentali sarà lo stesso perché la combinazione è guidata dalla piattaforma quindi la difficoltà risulta essere uguale qualunque sia la combinazione di servizi che l'utente deve chiedere.

L'obiettivo dell'indagine è, dunque, la valutazione della complessità di combinare servizi già esistenti.

Il caso di studio scelto dovrà essere indagato da una batteria di 15 soggetti sperimentali con diverse competenze, operanti con device di tipologia diversa (desktop e mobile) e selezionati tra le persone che potenzialmente potrebbero "vivere" il progetto I-Mole poiché impegnati in attività lavorative presso Assioma o presso le aziende del gruppo TXT, che si sono dette interessate alla validazione della piattaforma.

Ogni attività effettuata dai soggetti sperimentali dovrà essere dovrà monitorata da un tutor di Assioma e misurata; a tal fine le misure da rilevare sono due:

- una quantitativa: (TPR) il Tempo Persona impiegato, e dunque Rilevato, per la produzione del servizio da proporre a catalogo;
- una qualitativa derivante dal questionario CES (Customer Effort Score) che ogni soggetto sperimentale dovrà compilare dopo che avrà eseguito il compito assegnatogli.

Inoltre, come evidenziato in seguito, i 15 soggetti sperimentali dovranno compilare il questionario SUS, per la valutazione dell'usabilità della piattaforma I-Mole, e il questionario UEQ, per la valutazione della User eXperience della stessa piattaforma.

7.3 Specificazione dei servizi

L'obiettivo di questa macroarea dell'indagine è la validazione della complessità di specificare il servizio richiesto dal cliente nella "linea di servizi" che gli propone il sistema.

La validazione vedrà impegnati tutti i 45 utenti sperimentali individuati; infatti, essa prevederà l'esecuzione di tre task da far eseguire a 15 soggetti sperimentali estratti dal paniere dei 45 utenti individuati; ogni task rappresenterà un bisogno da soddisfare che dovrà essere illustrato alla relativa coorte per la corretta esecuzione dello stesso. Il tempo a disposizione del soggetto sperimentale non deve avere limiti predefiniti; inoltre, il soggetto sperimentale può decidere di smettere di eseguire l'azione nel momento in cui ritenesse di non riuscire a concluderla. In questo caso, l'azione deve essere considerata "fallita". Per questa indagine, un tutor, appositamente incaricato, dovrà valutare se il servizio rilevato da ognuno dei soggetti sperimentali è "corretto"/"errato" rispetto al bisogno assegnato.

L'obiettivo dell'indagine è la validazione della complessità di specificare il servizio richiesto dal cliente nella "linea di servizi" che gli propone il sistema. L'indagine ha previsto le seguenti azioni:

- si deve specificare il servizio che soddisfa il bisogno da soddisfare tra tutte le alternative che il sistema propone per ogni servizio. Anche in questo caso la difficolta della ricerca è diversa, dipendentemente dal bisogno che si deve soddisfare e che deve essere specificato in termini di caratteristiche del servizio cercato per soddisfarlo. Perciò, anche in questo caso si usano tre richieste diverse e ognuna di tali richieste è risolta da una coorte di almeno 15 persone per avere un buon risultato statistico.

Ogni attività effettuata dai soggetti sperimentali dovrà essere monitorata da un tutor di Assioma e misurata; a tal fine le misure da rilevare sono due:

- una quantitativa: (TPR) il Tempo Persona impiegato, e dunque Rilevato, per la produzione del servizio da proporre a catalogo;
- una qualitativa derivante dal questionario CES (Customer Effort Score) che ogni soggetto sperimentale dovrà compilare dopo che avrà eseguito il compito assegnatogli.

Inoltre, come evidenziato in seguito, i 45 soggetti sperimentali dovranno compilare il questionario SUS, per la valutazione dell'usabilità della piattaforma I-Mole, e il questionario UEQ, per la valutazione della User eXperience della stessa piattaforma.

7.4 Ricerca di un servizio

L'**obiettivo** di questa macroarea dell'indagine è la validazione della complessità della ricerca di un servizio richiesto dal cliente; ovvero, quanto la stessa risenta della complessità della piattaforma I-Mole.

L'indagine prevederà le seguenti azioni:

si deve cercare un servizio presente nel catalogo conoscendo le caratteristiche del servizio che si desidera, navigando sul sito web. In questo caso la difficoltà della ricerca è diversa per caratteristiche differenti; pertanto, si usano tre richieste diverse e ognuna di tali richieste è risolta da una coorte di almeno 15 persone per avere un buon risultato statistico.

La validazione vedrà impegnati tutti i 45 utenti sperimentali individuati; infatti, essa prevederà l'esecuzione di tre task da far eseguire a 15 soggetti sperimentali estratti dalla coorte dei 45 utenti individuati; ogni task rappresenterà un servizio da ricercare.

Ogni servizio dovrà essere illustrato alla relativa coorte per la corretta esecuzione dello stesso. Il tempo a disposizione del soggetto sperimentale non deve avere limiti predefiniti; inoltre, il soggetto sperimentale può decidere di smettere di eseguire l'azione nel momento in cui ritenesse di non riuscire a concluderla. In questo caso, l'azione deve essere considerata "fallita".

Per questa indagine, un tutor, appositamente incaricato, dovrà valutare se il servizio rilevato da ognuno dei soggetti sperimentali fosse "corretto" / "errato" rispetto al bisogno assegnato.

Ogni attività effettuata dai soggetti sperimentali dovrà essere dovrà monitorata da un tutor di Assioma e misurata; a tal fine le misure da rilevare sono due:

- una quantitativa: (TPR) il Tempo Persona impiegato, e dunque Rilevato, per la ricerca del servizio assegnato;
- una qualitativa derivante dal questionario CES (Customer Effort Score) che ogni soggetto sperimentale dovrà compilare dopo che avrà eseguito il compito assegnatogli.

Inoltre, come evidenziato in seguito, i 45 soggetti sperimentali dovranno compilare il questionario SUS, per la valutazione dell'usabilità della piattaforma I-Mole, e il questionario UEQ, per la valutazione della User eXperience della stessa piattaforma.

7.5 Usabilità ed UX del portale

L'**obiettivo** della sperimentazione è la validazione dell''usabilità della piattaforma e della User Experience.

I 45 soggetti sperimentali individuati per le precedenti indagini, dopo aver eseguito le azioni previste dai task a loro assegnati ed aver risposto al quesito per il calcolo dell'indice CES dovranno, infine, rispondere ai due questionari SUS ed UEQ previsti per la valutazione della Usabilità e della UX.

La compilazione di questi questionari sarà richiesta appena terminate le azioni precedenti per aumentare l'accuratezza delle risposte da parte dei soggetti sperimentali. Ovviamente, il processo di indagine sarà preventivamente spiegato ai soggetti sperimentali per avere la loro piena partecipazione e ridurre quando più possibile lo stress derivante dalla pluralità delle azioni richieste.

Dunque, l'indagine prevederà la seguente **azione**:

- ogni soggetto sperimentale dopo aver eseguito il task assegnato nei precedenti punti risponde ai due seguenti questionari:
  - Il System Usability Scale (SUS) (cfr TR4.3), per la **Usabilità;**
  - Il questionario User Experience Questionnaire (UEQ) (cfr TR4.3), per la **User eXperience**.

Dai risultati attesi potrebbe essere necessario modificare il portale per migliorare uno o più parametri di usabilità o di user experience. In tal caso, dopo ogni modifica si dovrà ripetere il sondaggio, si estrarranno i risultati sperimentali e si rifaranno i calcoli per ottenere i nuovi risultati.

L'efficacia dell'innovazione introdotta sarà provata rilevando le seguenti misure:

- Le rilevazioni del Sondaggio SUS effettuato da ogni soggetto sperimentale.
- Le rilevazioni dello User Experience Questionnaire (UEQ) effettuato da ogni soggetto sperimentale.

Nel progetto I-Mole al fine di condurre test di usabilità e test UX efficaci si è deciso di individuare una pluralità di profili utente, i soggetti sperimentali (cfr TR4.3), che riflettano gli utenti potenziali della piattaforma web implementata: questo approccio è fondamentale per ottenere risultati rappresentativi ed affidabili.

Quindi, soggetti sperimentali con competenze ed abilità diverse per ottenere feedback su come diverse categorie di utenti interagiscono con la piattaforma I-Mole. Persone con background diversi, età, livelli di istruzione, culture e contesti di utilizzo. I più esperti possono utilizzare funzionalità avanzate,

mentre altri potrebbero essere nuovi e concentrarsi sulle caratteristiche di base. I diversi utenti possono rilevare problemi distinti: un utente esperto potrebbe notare questioni di efficienza o avanzate, mentre un utente meno esperto potrebbe sottolineare problemi di comprensibilità o accessibilità.

Anche la numerosità del campione dei soggetti sperimentali coinvolti nell'indagine è importante ed aumenta la validità statistica; i test di usabilità mirano a fornire informazioni rappresentative del pubblico target, e un campione più ampio riduce la possibilità di risultati influenzati da variazioni casuali o outliers.

Fatte tali considerazioni, nel progetto I-Mole è stato definito il numero di 45 soggetti sperimentali; questi ultimi sono opportunamente raggruppati in 6 classi, ovvero: principiante, avviato, esperto, mobile principiante, mobile avviato, mobile esperto (cfr. TR4.3) sulla base delle informazioni a disposizione della gestione "risorse umane" delle imprese di provenienza di ciascun soggetto sperimentale.

Inoltre, è il caso di evidenziare che tutti i soggetti sperimentali si sono candidati volontariamente dopo aver avuto un colloquio motivazionale ed informativo con gli sperimentatori.

Alla fine della sperimentazione, dunque, saranno raccolti i seguenti sondaggi:

- 15 sondaggi SUS per la valutazione dell'usabilità della Composizione dei Servizi
- 45 sondaggi SUS per la valutazione dell'usabilità della Specificazione dei Servizi
- 45 sondaggi SUS per la valutazione dell'usabilità della Ricerca dei Servizi
- 15 sondaggi UEQ per la valutazione della User eXperience della Composizione dei Servizi

- 45 sondaggi UEQ per la valutazione della User eXperience della Specificazione dei Servizi
- 45 sondaggi UEQ per la valutazione della User eXperience della Ricerca dei Servizi

Tale pluralità di sondaggi si rende necessaria in quanto è indispensabile che gli stessi siano compilati subito dopo l'esecuzione del task assegnato al Soggetto Sperimentale.

Tra l'altro è di tutta evidenza che ogni soggetto sperimentale compilerà come minimo due questionari SUS e UEQ. Dall'analisi di tali questionari si potranno ottenere utili indicazioni circa la bontà dell'interfaccia utente della piattaforma I-Mole.

7.6 Penetration test

La sperimentazione che Unisannio svolgerà sarà finalizzata a valutare il grado di robustezza del sistema sviluppato rispetto alle proprietà di sicurezza, ovvero integrità, disponibilità e confidenzialità della piattaforma Imole.

I risultati della sperimentazione, però, non saranno utilizzati solo per acquisire una fotografia dello stato di sicurezza della piattaforma, ma avranno una ricaduta molto più ampia.

Essi saranno utili a fornire un feedback per migliorare il processo di progettazione, sviluppo e testing, identificando quali sono le maggiori debolezze nella realizzazione di sistemi nella prospettiva della sicurezza.

Tali feedback potranno identificare:

- carenze nel monitoraggio delle vulnerabilità che affliggono le tecnologie utilizzate nello stack software;
- cattiva identificazione, selezione, progettazione o configurazione di controlli sviluppati per garantire le proprietà di sicurezza del sistema;
- pattern di codice errati che possono esporre ad un rischio di violazione delle politiche di riservatezza e sicurezza del sistema o dei dati da esso gestiti o utilizzati.

A valle dell'analisi dei risultati, sarà possibile valutare:

- le capacità e le competenze per l'implementazione dei paradigmi di progettazione, sviluppo e testing che garantiscono la built-in security;
- la completezza e la correttezza dei processi che devono essere acquisiti ed implementati per garantire la built-in security all'interno dei prodotti;
- l'adozione corretta e completa degli strumenti che consentono la realizzazione di prodotti intrinsecamente sicuri, quali analizzatori statici di codice, asset management, scanner di vulnerabilità, tool di supporto al test di sicurezza;
- l'adozione corretta e completa delle pratiche di sviluppo che consentono la realizzazione di codice sicuro, come framework e standard per lo sviluppo sicuro, framework e standard per il test di sicurezza, checklist per la verifica della sicurezza del codice, input filtering, autenticazione, protezione dei canali di comunicazione, attuazione del principio need to know, applicazione del principio di completa mediazione.

Le domande di ricerca che hanno orientato la sperimentazione sono di seguito dettagliate.

RQ1: Qual è la copertura delle vulnerabilità riscontrate dai vulnerability scanner?

Questa RQ mira a valutare la capacità dei vulnerability scanner di coprire le diverse classi di vulnerabilità. Purtroppo, a causa della rapida proliferazione delle vulnerabilità, non sempre i vulnerability scanner sono in grado di aggiornarsi prontamente per tenerne conto. Questo fenomeno causa inefficienze sporadiche, sottolineando l'importanza di avere una mappatura dettagliata delle vulnerabilità che ciascuno scanner è in grado di rilevare. Le vulnerabilità saranno categorizzate seguendo lo schema delle CWE fornito dal MITRE nel Common Vulnerability and Exposure Database.

RQ2: Qual è l'obsolescenza delle vulnerabilità riscontrate dai vulnerability scanner?

Questa RQ si propone di valutare l'obsolescenza delle vulnerabilità individuate nel prodotto. L'obsolescenza rappresenta un parametro di grande rilevanza, in quanto è direttamente correlato alla finestra temporale durante

la quale gli attaccanti possono sfruttare le vulnerabilità. Maggiore è questa finestra, maggiore è il rischio di essere soggetti a un numero crescente di attacchi, aumentando così l'esposizione al pericolo. Tuttavia, l'obsolescenza non è l'unico elemento che misura tale esposizione; verranno quindi valutati altri parametri, tra cui il vettore di attacco impiegato, spesso indicativo della pericolosità della vulnerabilità, e il CVSS, un indice che misura l'impatto della stessa.

RQ3: Qual è la copertura tecnologica dei vulnerability scanner?

Questa RQ si propone di valutare l'ampiezza delle tecnologie considerate dai vulnerability scanner. Purtroppo, questi scanner non riescono a coprire tutte le tecnologie esistenti, dato che una copertura totale sarebbe impraticabile. Inoltre, anche se uno scanner dovesse coprire una determinata tecnologia, potrebbe non garantire lo stesso livello di efficacia. Pertanto, è cruciale realizzare un benchmark anche in questo contesto, considerando la diversità delle tecnologie e valutando la capacità degli scanner di affrontarle in modo efficace.

7.7 Sicurezza con la blockchain

La sicurezza è un aspetto fondamentale da tenere in considerazione quando si elaborano dati, si creano asset monetari, si scambiano monete. Per verificare l'avvenuta transazione, le informazioni relative a venditore, acquirente e id dell'ordine sono state inserita in una blockchain, in modo che tali informazioni risultino salvate e immutabili nel tempo.

In questa sezione è presente un'introduzione su Ethereum e viene descritto come è stata configurata la rete di nodi da utilizzare per la sperimentazione.

7.7.1 Ethereum

Con la piattaforma blockchain decentralizzata Ethereum viene stabilita una rete peer-to-peer per eseguire e verificare in maniera sicura uno smart contract. Uno smart contract è un programma eseguito sulla blockchain di

Ethereum costituito da codice (le funzioni) e dati (lo stato)[1]. Attraverso gli smart contract, i partecipanti possono scambiare transazioni senza la necessità di un'autorità centrale attendibile. I record delle transazioni sono immutabili, verificabili e distribuiti in modo sicuro su tutta la rete, concedendo ai partecipanti la piena proprietà e visibilità dei dati delle transazioni. Le transazioni vengono inviate e ricevute da account Ethereum creati dagli utenti. Un mittente firma transazioni e spende Ether, la criptovaluta nativa di Ethereum, per il costo di elaborazione delle transazioni sulla rete.[2]

7.7.2 Rete Blockchain Ethereum privata

In questa sezione è descritto come è stato configurato ed eseguito un cluster di 8 nodi per creare una rete Ethereum privata in ambiente Docker.[3]

7.7.2.1 I nodi della rete Ethereum privata

La rete Ethereum è una rete peer-to-peer formata da più nodi che eseguono il client Ethereum mediante Geth.

In questo caso, la rete privata viene configurata come segue:

1. **Bootnode** : il nodo bootstrap utilizzato per rilevare i peer. Rimane in ascolto sulla porta *30303*, gli altri nodi (7) che si uniscono alla rete si connettono prima a questo bootnode.
2. **Endpoint JSON-RPC** : questo nodo espone l'API JSON-RPC sull'endpoint HTTP sulla porta *8545*. Attraverso tale porta è possibile l'interazione esterna con questa blockchain privata.

[1] https://ethereum.org/it/developers/docs/smart-contracts
[2] https://aws.amazon.com/it/blockchain/what-is-ethereum/#:~:text=Ethereum%20%C3%A8%20una%20piattaforma%20blockchain,senza%20un'autorit%C3%A0%20centrale%20attendibile.
[3] https://medium.com/scb-digital/running-a-private-ethereum-blockchain-using-docker-589c8e6a4fe8

3. **Miner** : questo nodo si occupa del mining (il processo di creazione di un nuovo blocco nella blockchain). Quando il nodo miner estrae con successo un nuovo blocco, riceve i premi nell'account configurato.

Per i nodi Ethereum nella blockchain privata, è utilizzato Go Ethereum (Geth) come client. È stato creato un Dockerfile per costruire l'immagine del client Ethereum.

Come descritto in precedenza, la rete Ethereum privata è composta da 8 nodi collegati tra loro. Per collegarli, tutti i nodi sono inseriti nella sottorete 172.16.254.0/28 in una rete bridge Docker.

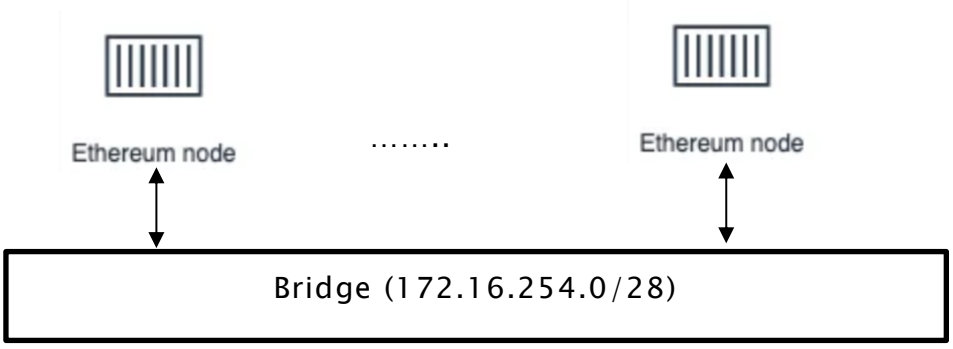

Figura 7.1 Il cluster di nodi Ethereum in Docker

Poiché si vogliono avere 3 nodi (bootnode, json-RPC e miner) che svolgono ruoli diversi nella blockchain privata, ciascun nodo sarà configurato con un diverso set di parametri, mentre i restanti nodi avranno la stessa configurazione.

8. BIBLIOGRAFIA

Amschler Andrews A, Pradhan AS (2001) Ethical issues in empirical software engineering: the limits of policy. Empir Softw Eng 6(2):105–110

Anastas JW, MacDonald ML (1994) Research design for the social work and the human services. New York, Lexington

Andersson C, Runeson P (2007a) A replicated quantitative analysis of fault distribution in complex software systems. IEEE Trans Softw Eng 33(5):273–286

Andersson C, Runeson P (2007b) A spiral process model for case studies on software quality monitoring—method and metrics. Softw Process Improv Pract 12(2):125–140

Avison D, Baskerville R, Myers M (2001) Controlling action research projects. Inf Technol People 14(1):28–45

Basili VR, Weiss DM (1984) A methodology for collecting valid software engineering data. IEEE Trans Softw Eng SE10(6):728–739

Basili VR, Selby RW, Hutchens DH (1986) Experimentation in Software Engineering. IEEE Trans Softw Eng SE12(7):733–744

Baskerville RL, Wood-Harper AT (1996) A critical perspective on action research as method for information systems research. J Inf Technol 11:235–246

Benbasat I, Goldstein DK, Mead M (1987) The case research strategy in studies of information systems. MIS Q 11(3):369–386

Corbin J, Strauss C (2008) Basics of qualitative research, 3rd edn. Sage

Dalkey N, Helmer O (1963) An experimental application of the delphi method to the use of experts. Manage Sci 9(3):458–467

Dittrich Y (ed) (2007) Special issue on qualitative software engineering research. Inf Softw Technol 49(6):531–694.

Dittrich Y, Rönkkö K, Eriksson J, Hansson C, Lindeberg O (2008) Cooperative method development. combining qualitative empirical research with method, technique and process improvement. Empir Softw Eng 13(3):231–260

H. Dzwigol, (2020), Methodological and empirical platform of triangulation in strategic management, Academy of strategic management journal,volume 19, issue 4, 2020

Eisenhardt KM (1989) Building theories form case study research. Acad Manage Rev 14(4):532–550

Easterbrook S, Singer J, Storey M-A, Damian D (2008) Selecting empirical methods for software engineering research, Chapter 11 in Shull et al. (2008

Fenton N, Pfleeger SG (1996) *Software Metrics—A Rigorous and Practical Approach*, Thomson Computer

Flynn BB, Sakakibara S, Schroeder RG, Bates K, Flynn EJ (1990) Empirical research methods in operations management. Oper Manage 9(2):250–284

Flyvbjerg B (2007) Five misunderstandings about case-study research. In *Qualitative Research Practice: Concise Paperback Edition*. Sage, pp 390–404

Gorschek T, Garre P, Larsson S, Wohlin C (2006) A model for technology transfer in practice. IEEE Softw 23(6):88–95

Hannay JE, Sjøberg DIK, Dybå TA (2007) Systematic review of theory use in software engineering experiments. IEEE Trans Softw Eng 33(2):87–107

Hove SE, Anda BCD (2005) Experiences from conducting semi-structured interviews in empirical software engineering research. Proceeding's 11th IEEE International Software Metrics Symposium (Metrics 2005) 23:1–10

Höst M, Runeson P (2007) Checklists for Software Engineering Case Study Research, In *Proceedings First International Symposium on Empirical Software Engineering and Measurement*, pp 479–481

Iversen JH, Mathiassen L, Nielsen PA (2004) Managing risk in software process improvement: an action research approach. MIS Q 28(3):395–433

Jedlitschka A, Pfahl D (2005) Reporting guidelines for controlled experiments in software engineering, In *Proceedings of ACM/IEEE International Symposium on Empirical Software Engineering*, pp 95–104, see also Chapter 8 in Shull et al. (2008)

Johnson P, Kou H, Paulding M, Zhang Q, Kagawa A, Yamashita T (2005) Improving software development management through software project telemetry. IEEE Softw 22(4):76–85

Karahasanović A, Anda B, Arisholm E, Hove SE, Jørgensen M, Sjøberg DIK, Welland R (2005) Collecting feedback during software engineering experiments. Empir Softw Eng 10:113–147

Karlström D (2004) *Integrating Management and Engineering Processes in Software Product Development*, PhD Thesis ISRN LUTEDX/TETS—1069–SE+230p, Lund University.

Karlström D, Runeson P (2005) Combining agile methods with stage-gate project management. IEEE Softw 22(3):43–49

Karlström D, Runeson P (2006) Integrating agile software development into stage-gate product development. Empir Softw Eng 11:203–225

Klein HK, Myers MD (1999) A set of principles for conducting and evaluating interpretative field studies in information systems. MIS Q 23(1):67–

Kitchenham B, Pickard L, Pfleeger SL (1995) Case studies for method and tool evaluation. IEEE Softw 4(12):52–62

Kitchenham B, Pfleeger SM, Pickard LM, Jones PW, Hoaglin DC, El Eman K, Rosenberg J (2002) Preliminary guidelines for empirical research in software engineering. IEEE Trans Softw Eng 28(8):721–734

Kitchenham B (2007) *Guidelines for performing Systematic Literature Reviews in Software Engineering*, Version 2.3, EBSE Technical Report EBSE–2007–01, Keele University and University of Durham

Kitchenham B, Al-Khilidar H, Ali Babar M, Berry M, Cox K, Keung J, Kurniawati F, Staples M, Zhang H, Zhu L (2008) Evaluating guidelines for reporting empirical software engineering studies. Empir Softw Eng 13(1):97–121

Lee AS (1989) A scientific methodology for MIS case studies. MIS Q 13(1):33–54

Lethbridge TC, Sim SE, Singer J (2005) Studying software engineers: data collection techniques for software field studies. Empir Softw Eng 10(3):311–341 see also Chapter 1 in Shull et al. (2008)

Moher T, Schneider GM (1981) Methods for improving controlled experimentation in software engineering, *Proceedings of the 5th International Conference on Software Engineering*

Owen S, Budgen D, Brereton P (2006) Protocol analysis: a neglected practice. Commun ACM 49(2):117–

Perry DE, Sim SE, Easterbrook S (2005) Case studies for software engineers, *29th Annual IEEE/NASA Software Engineering Workshop—Tutorial Notes* pp 96–159

Pervan G, Maimbo H (2005) Designing a case study protocol for application in IS research, *The Ninth Pacific Conference on Information Systems* pp 1281–1292

Ramesh V, Glass RL, Vessey I (2004) Research in computer science: an empirical study. J Syst Softw 70(1–2):165–176

Regnell B, Höst M, Natt och Dag J, Beremark P, Hjelm T (2001) An industrial case study on distributed prioritisation in market-driven requirements engineering for packaged software. Requirements Eng

Robinson H, Segal J, Sharp H (2007) Ethnographically informed empirical studies of software practice. Inf Softw Technol 49:540–551

Robson C (2002) *Real World Research*. Blackwell, (2nd edition)

Seaman C (1999) Qualitative methods in empirical studies of software engineering. IEEE Trans Softw Eng 25(4):557–572 see also Chapter 2 in Shull et al. (2008)

Sharp H, Robinson H (2004) An ethnographic study of XP practice. Empir Softw Eng 9(4):353–375

Shull F, Feldman RL (2008) Building theories from multiple evidence sources. In: Shull F et al (ed) Guide to advanced empirical software engineering. Springer-Verlag, London

Shull F, Basili V, Carver J, Maldonado JC, Travassos GH, Mendonca M, Fabbri S (2002) Replicating software engineering experiments: addressing the tacit knowledge problem, *Proceedings on International Symposium Empirical Software Engineering* pp 7–16

Shull F, Singer J, Sjøberg D (eds) (2008) Guide to Advanced Empirical Software Engineering. Springer-Verlag: London

Sim SE, Singer J, Storey M-A (2001) Beg, borrow, or steal: using multidisciplinary approaches in empirical software engineering research, an ICSE 2000 workshop report. Empir Softw Eng 6(1):85–93

Singer J, Vinson NG (2002) Ethical issues in empirical studies of software engineering. IEEE Trans Softw Eng 28(12):1171–1180

Sjøberg DIK, Dybå T, Anda BCD, Hannay J (2008) Building theories in software engineering. In: Shull F et al (ed) Guide to advanced empirical software engineering. Springer-Verlag, London

Sjøberg DIK, Hannay JE, Hansen O, Kampenes VB (2005) A survey of controlled experiments in software engineering. IEEE Trans Softw Eng 31(9):733–753

Stake RE (1995) *The art of case study research*. Sage

Tichy WF (1998) Should computer scientists experiment more? Computer 31(5):32–40

Tichy WF (2000) Hints for reviewing empirical work in software engineering. Empir Softw Eng 5(4):309–

van Solingen R, Berghout E (1999) *The goal/question/metric method. A practical guide for quality improvement of software development*. McGraw-Hill

von Mayrhauser A, Vans AM (1996) Identification of dynamic comprehension processes during large scale maintenance. IEEE Trans Softw Eng 22(6):424–438

Wallace C, Cook C, Summet J, Burnett M (2002) Human centric computing languages and environments. *Proceeding Symposia on Human Centric Computing Languages and Environments*

Wohlin C, Höst M (2001) Special section: controlled experiments in software engineering, guest editorial. Inf Softw Technol 43(15):921–924

Wohlin C, Höst M, Ohlsson MC, Regnell B, Runeson P, Wesslén A (2000) *Experimentation in software engineering—an introduction*. Kluwer

Wohlin C, Höst M, Henningsson K (2003) Empirical research methods in software engineering. In: Conradi R, Wang AI (eds) *Empirical Methods and Studies in Software Engineering—Experiences from ESERNET*, Springer

Yin RK (2003) Case study research. Design and methods, 3rd edn. London, Sage

Zelkowitz MV, Wallace RW (1998) Experimental models for validating technology. IEEE Comput 31(5):23-31