In [31]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt


data = pd.read_csv('cgwb-changes-in-depth-to-water-level.csv')

print("Dataset shape:", data.shape)
print("Column names:")
print(data.columns.tolist())
print("First few rows:")
display(data[data['state_name'] == 'Andhra Pradesh'])
info = data[data['state_name'] == 'Andhra Pradesh']
print("**********Andhra Pradesh**********")
display(info.head())
```

```
Dataset shape: (550850, 14)
Column names:
['id', 'date', 'state_name', 'state_code', 'district_name', 'district_code', 'station_name', 'latitude', 'longit
ude', 'basin', 'sub_basin', 'source', 'currentlevel', 'level_diff']
First few rows:
```

	id	date	state_name	state_code	district_name	district_code	station_name	latitude	longitude	basin	sub_basin
167286	167286	2013-01-04	Andhra Pradesh	28	Krishna	510	Gopalapuram	16.97917	80.65833	Krishna	Krishna Lowe
167287	167287	2013-05-14	Andhra Pradesh	28	Krishna	510	Gopalapuram	16.97917	80.65833	Krishna	Krishna Lowe
167288	167288	2013-01-04	Andhra Pradesh	28	Krishna	510	Gudivada	16.43333	80.99167	East Flowing Rivers Between Mahanadi And Pennar	Eas Flowing Rivers Between Godavar And Krishna
167289	167289	2013-05-14	Andhra Pradesh	28	Krishna	510	Gudivada	16.43333	80.99167	East Flowing Rivers Between Mahanadi And Pennar	Eas Flowing Rivers Between Godavar And Krishna
167290	167290	2014-01-04	Andhra Pradesh	28	Krishna	510	Gudivada	16.43333	80.99167	East Flowing Rivers Between Mahanadi And Pennar	Eas Flowing Rivers Between Godavar And Krishna
...	...	...	...	...	...	...	...	...	...	...	..
189210	189210	2019-08-19	Andhra Pradesh	28	Kurnool	511	Rudravaram-DW13	15.85028	78.47083	Krishna	Krishna Middle
189211	189211	2019-10-31	Andhra Pradesh	28	Kurnool	511	Rudravaram-DW13	15.85028	78.47083	Krishna	Krishna Middle
189212	189212	2020-01-09	Andhra Pradesh	28	Kurnool	511	Rudravaram-DW13	15.85028	78.47083	Krishna	Krishna Middle
189213	189213	2020-11-29	Andhra Pradesh	28	Kurnool	511	Rudravaram-DW13	15.85028	78.47083	Krishna	Krishna Middle
189214	189214	2021-01-09	Andhra Pradesh	28	Kurnool	511	Rudravaram-DW13	15.85028	78.47083	Krishna	Krishna Middle

21929 rows × 14 columns

```
**********Andhra Pradesh**********
```

	id	date	state_name	state_code	district_name	district_code	station_name	latitude	longitude	basin	sub_basin
167286	167286	2013-01-04	Andhra Pradesh	28	Krishna	510	Gopalapuram	16.97917	80.65833	Krishna	Krishna Lowe
167287	167287	2013-05-14	Andhra Pradesh	28	Krishna	510	Gopalapuram	16.97917	80.65833	Krishna	Krishna Lowe
167288	167288	2013-01-04	Andhra Pradesh	28	Krishna	510	Gudivada	16.43333	80.99167	East Flowing Rivers Between Mahanadi And Pennar	East Flowing Rivers Between Godavar And Krishna
167289	167289	2013-05-14	Andhra Pradesh	28	Krishna	510	Gudivada	16.43333	80.99167	East Flowing Rivers Between Mahanadi And Pennar	East Flowing Rivers Between Godavar And Krishna
167290	167290	2014-01-04	Andhra Pradesh	28	Krishna	510	Gudivada	16.43333	80.99167	East Flowing Rivers Between Mahanadi And Pennar	East Flowing Rivers Between Godavar And Krishna

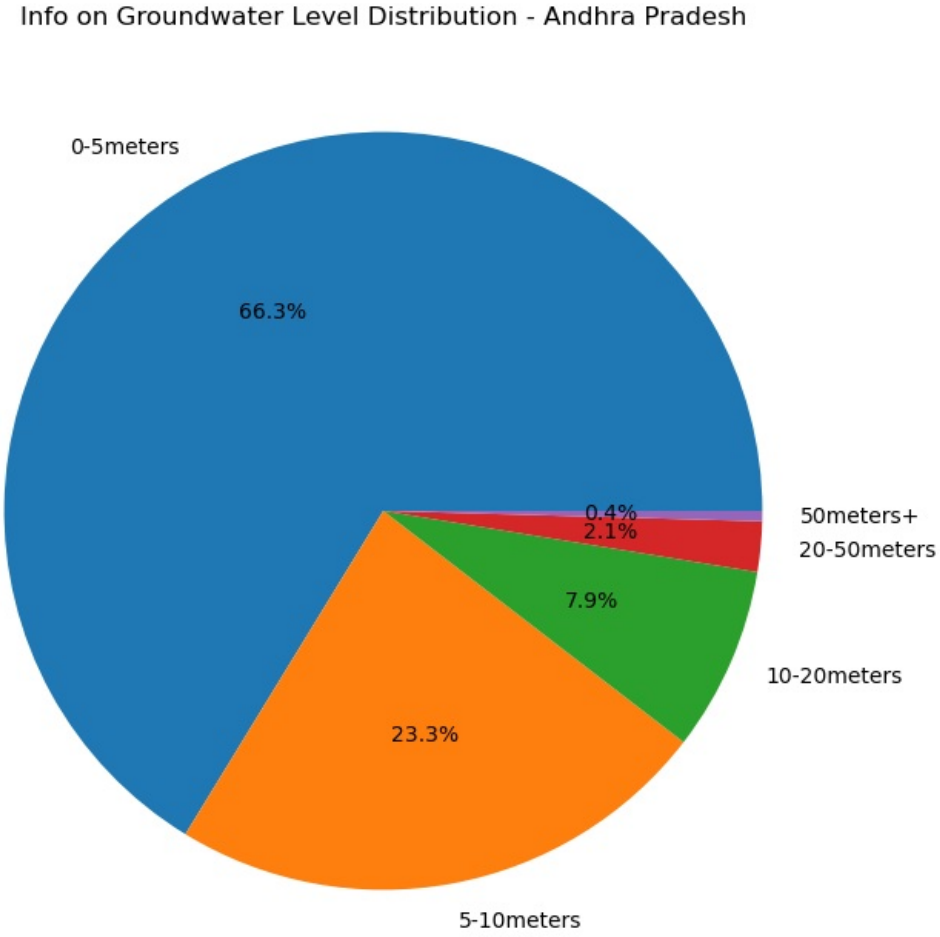
In [32]:
```
print("Data types:")
print(data.dtypes)
print("Finding Missing values:")
print(data.isnull().sum())
print("Basic statistics for numerical columns:")
print(data[['currentlevel', 'level_diff']].describe())


data['date'] = pd.to_datetime(data['date'])
print("\
Date range:", data['date'].min(), "to", data['date'].max())
print("Number of unique stations:", data['station_name'].nunique())
```

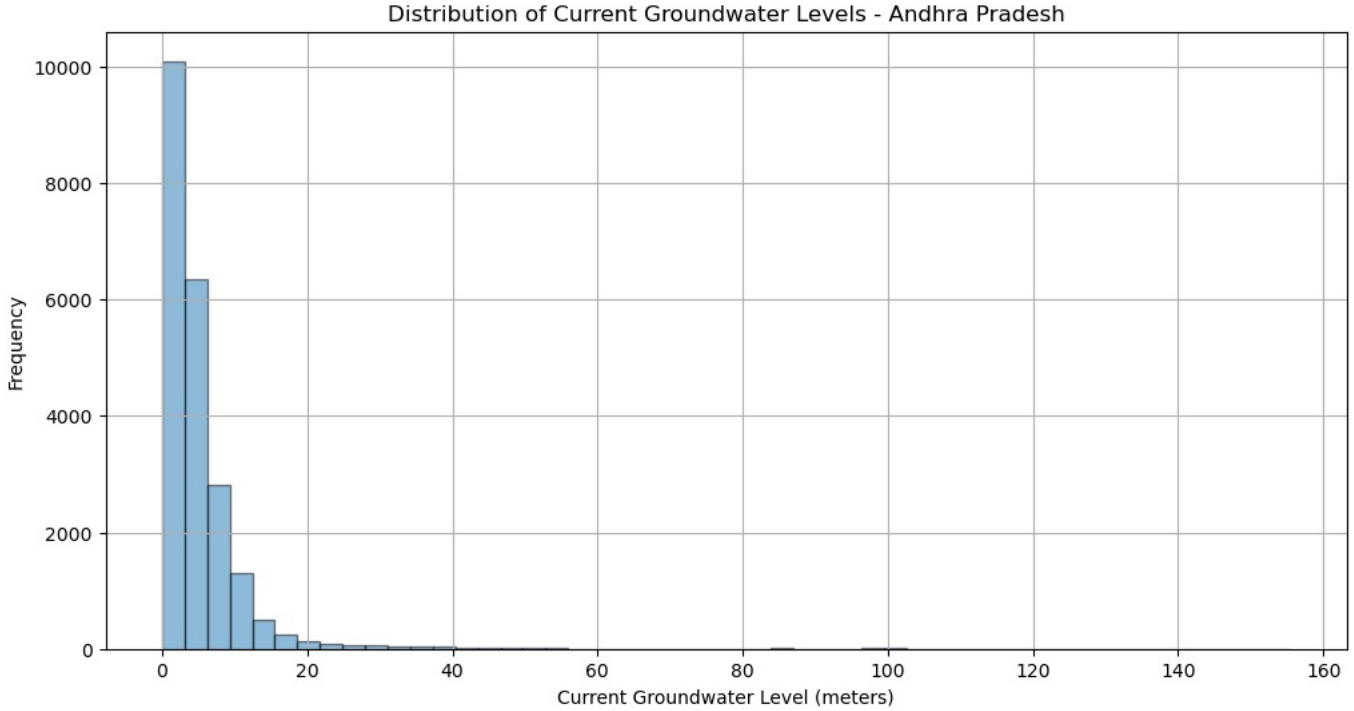
```
Data types:
id                 int64
date              object
state_name        object
state_code         int64
district_name     object
district_code      int64
station_name      object
latitude         float64
longitude        float64
basin             object
sub_basin         object
source            object
currentlevel     float64
level_diff       float64
dtype: object
Finding Missing values:
id               0
date             0
state_name       0
state_code       0
district_name    0
district_code    0
station_name     0
latitude         0
longitude        0
basin            0
sub_basin        0
source           0
currentlevel     0
level_diff       0
dtype: int64
Basic statistics for numerical columns:
        currentlevel     level_diff
count  550850.000000  550850.000000
mean        8.515535       0.015445
std        10.717358       5.773125
min         0.000000    -276.580000
25%         2.900000      -1.300000
50%         5.450000       0.300000
75%         9.650000       1.550000
max       299.300000     253.510000
Date range: 2013-01-01 00:00:00 to 2023-03-31 00:00:00
Number of unique stations: 23078
```

In [33]:

```
andhra_data = data[data['state_name'] == 'Andhra Pradesh'].copy()

print("Andhra Pradesh data shape:", andhra_data.shape)
print("Current level statistics for Andhra Pradesh:")
print(andhra_data['currentlevel'].describe())

level_ranges = pd.cut(andhra_data['currentlevel'],
                      bins=[0, 5, 10, 20, 50, float('inf')],
                      labels=['0-5meters', '5-10meters', '10-20meters', '20-50meters', '50meters+'])

range_counts = level_ranges.value_counts()

plt.figure(figsize=(10, 8))
plt.pie(range_counts.values, labels=range_counts.index, autopct='%1.1f%%')
plt.title("Info on Groundwater Level Distribution - Andhra Pradesh")
plt.show()

plt.figure(figsize=(12, 6))
plt.hist(andhra_data['currentlevel'], bins=50, alpha=0.5, edgecolor='black')
plt.title("Distribution of Current Groundwater Levels - Andhra Pradesh")
plt.xlabel("Current Groundwater Level (meters)")
plt.ylabel("Frequency")
plt.grid()
plt.show()
```

```
Andhra Pradesh data shape: (21929, 14)
Current level statistics for Andhra Pradesh:
count    21929.000000
mean         5.237814
std          7.373601
min          0.000000
25%          1.820000
50%          3.410000
75%          6.210000
max        155.420000
Name: currentlevel, dtype: float64
```





In [34]:
```
info1 = display(data[data['district_name'] == 'Krishna'])
```

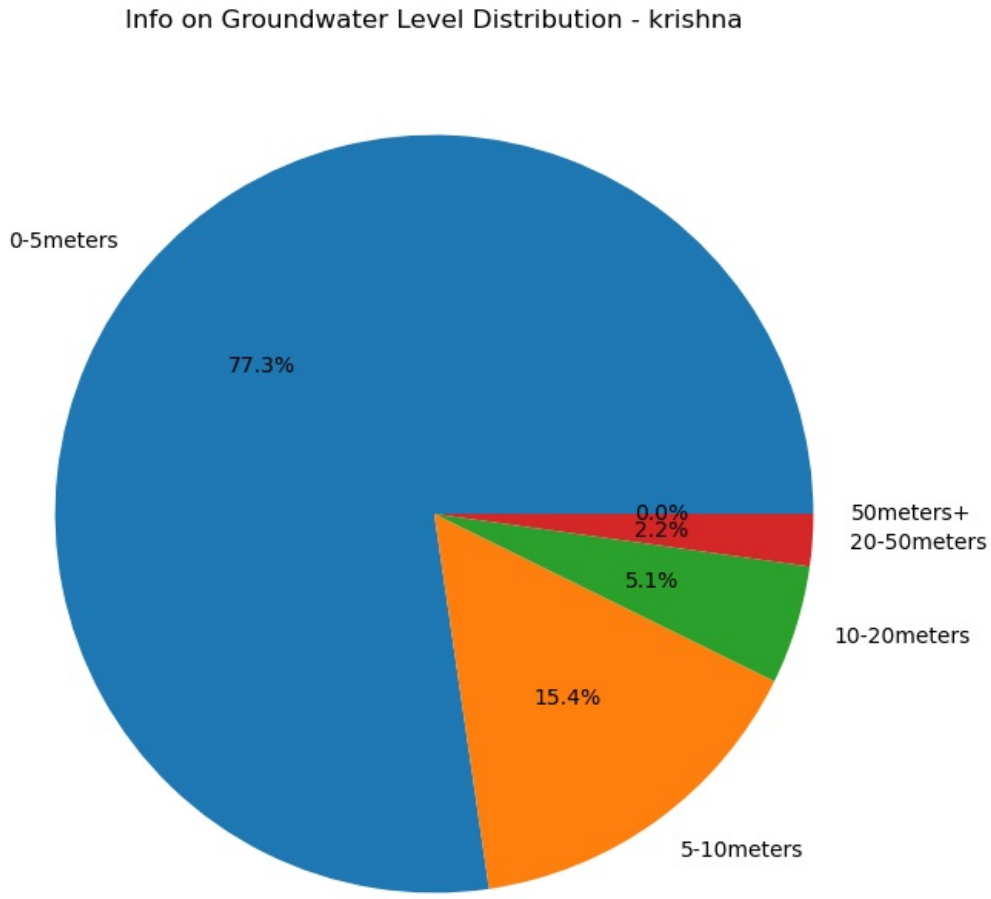
	id	date	state_name	state_code	district_name	district_code	station_name	latitude	longitude	basin	sub_basin
167286	167286	2013-01-04	Andhra Pradesh	28	Krishna	510	Gopalapuram	16.97917	80.65833	Krishna	Krishna Lowe
167287	167287	2013-05-14	Andhra Pradesh	28	Krishna	510	Gopalapuram	16.97917	80.65833	Krishna	Krishna Lowe
167288	167288	2013-01-04	Andhra Pradesh	28	Krishna	510	Gudivada	16.43333	80.99167	East Flowing Rivers Between Mahanadi And Pennar	Eas Flowing Rivers Between Godavar And Krishna
167289	167289	2013-05-14	Andhra Pradesh	28	Krishna	510	Gudivada	16.43333	80.99167	East Flowing Rivers Between Mahanadi And Pennar	Eas Flowing Rivers Between Godavar And Krishna
167290	167290	2014-01-04	Andhra Pradesh	28	Krishna	510	Gudivada	16.43333	80.99167	East Flowing Rivers Between Mahanadi And Pennar	Eas Flowing Rivers Between Godavar And Krishna
...	...	...	...	...	...	...	...	...	...	...	..
169176	169176	2019-08-19	Andhra Pradesh	28	Krishna	510	Kaikaluru1	16.55000	81.21667	East Flowing Rivers Between Mahanadi And Pennar	Eas Flowing Rivers Between Godavar And Krishna
169177	169177	2019-10-31	Andhra Pradesh	28	Krishna	510	Kaikaluru1	16.55000	81.21667	East Flowing Rivers Between Mahanadi And Pennar	Eas Flowing Rivers Between Godavar And Krishna
169178	169178	2020-01-09	Andhra Pradesh	28	Krishna	510	Kaikaluru1	16.55000	81.21667	East Flowing Rivers Between Mahanadi And Pennar	Eas Flowing Rivers Between Godavar And Krishna
169179	169179	2020-11-29	Andhra Pradesh	28	Krishna	510	Kaikaluru1	16.55000	81.21667	East Flowing Rivers Between Mahanadi And Pennar	Eas Flowing Rivers Between Godavar And Krishna
169180	169180	2021-01-09	Andhra Pradesh	28	Krishna	510	Kaikaluru1	16.55000	81.21667	East Flowing Rivers Between Mahanadi And Pennar	Eas Flowing Rivers Between Godavar And Krishna

1895 rows × 14 columns

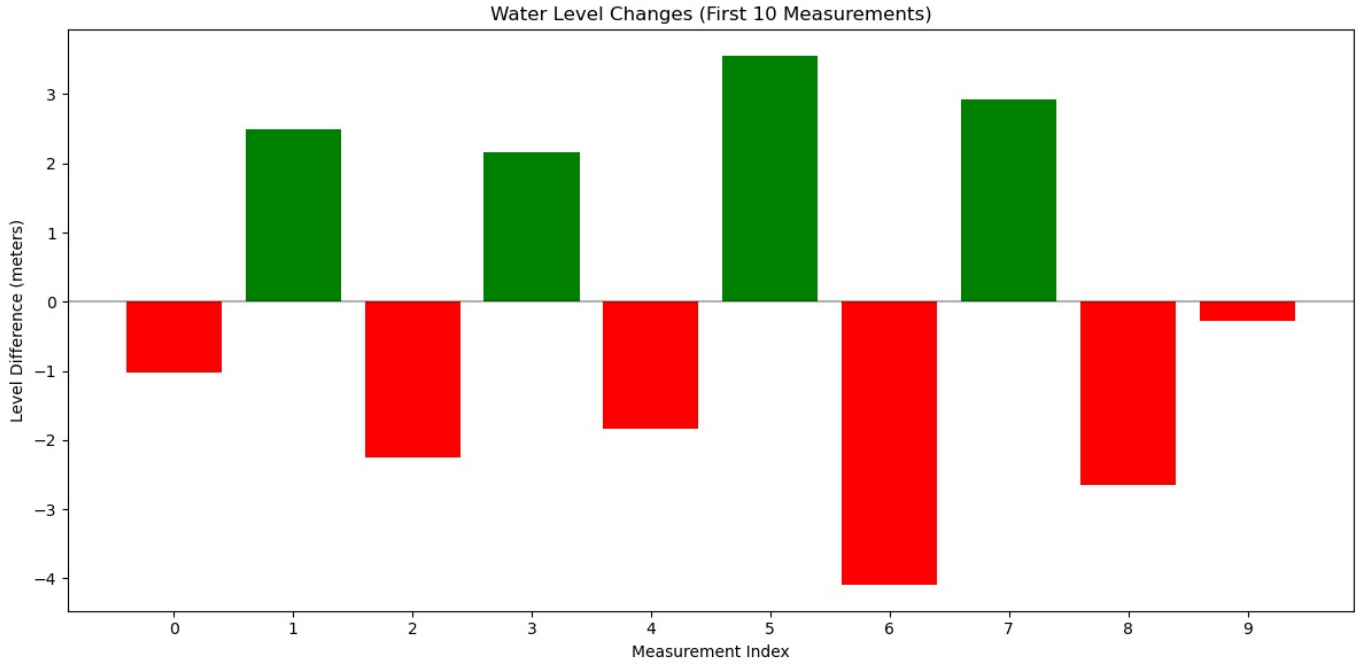
In [35]:
```
andhra_data1 = data[data['district_name'] == 'Krishna'].copy()
```

In [36]:
```
print("Andhra Pradesh data shape:", andhra_data1.shape)
print("Current level statistics for Andhra Pradesh:")
print(andhra_data1['currentlevel'].describe())

level_ranges = pd.cut(andhra_data1['currentlevel'],
                      bins=[0, 5, 10, 20, 50, float('inf')],
                      labels=['0-5meters', '5-10meters', '10-20meters', '20-50meters', '50meters+'])
range_counts = level_ranges.value_counts()

plt.figure(figsize=(10, 8))
plt.pie(range_counts.values, labels=range_counts.index, autopct='%1.1f%%')
plt.title("Info on Groundwater Level Distribution - krishna")
plt.show()
```

```
Andhra Pradesh data shape: (1895, 14)
Current level statistics for Andhra Pradesh:
count    1895.000000
mean        4.099636
std         4.816705
min         0.010000
25%         1.520000
50%         2.620000
75%         4.700000
max        38.750000
Name: currentlevel, dtype: float64
```

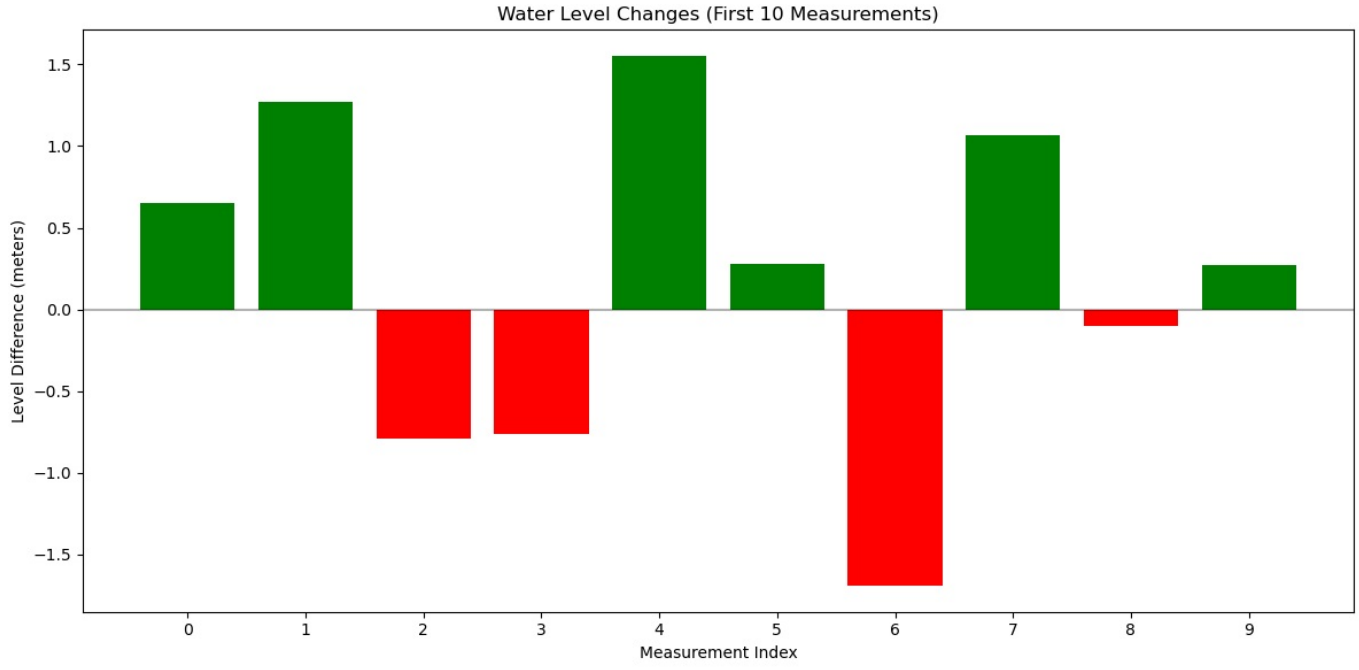


In [37]:
```python
level = pd.to_numeric(data['level_diff'])


valid_data = data[level.notna()].copy()
valid_level_diff = level[level.notna()]


colors = (valid_level_diff > 0.0).map({True: 'green', False: 'red'})

plt.figure(figsize=(12, 6))
plt.bar(range(10), valid_level_diff.head(10), color=colors.head(10))
plt.title("Water Level Changes (First 10 Measurements)")
plt.xlabel("Measurement Index")
plt.ylabel("Level Difference (meters)")
plt.axhline(y=0, color='black', linestyle='-', alpha=0.3)
plt.xticks(range(10))
plt.tight_layout()
plt.show()

print("Summary of level differences:")
print("Positive changes (rising water):", (valid_level_diff > 0).sum())
print("Negative changes (falling water):", (valid_level_diff < 0).sum())
print("No change:", (valid_level_diff == 0).sum())
```




```
Summary of level differences:
Positive changes (rising water): 317518
Negative changes (falling water): 233332
No change: 0
```

In [38]:
```
level = pd.to_numeric(data['level_diff'])


valid_data = data[level.notna()].copy()
valid_level_diff = level[level.notna()]


colors = (valid_level_diff > 0.0).map({True: 'green', False: 'red'})

plt.figure(figsize=(12, 6))
plt.bar(range(10), valid_level_diff.tail(10), color=colors.tail(10))
plt.title("Water Level Changes (First 10 Measurements)")
plt.xlabel("Measurement Index")
plt.ylabel("Level Difference (meters)")
plt.axhline(y=0, color='black', linestyle='-', alpha=0.3)
plt.xticks(range(10))
plt.tight_layout()
plt.show()
```



In [ ]:

In [ ]: